



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

**A SIMULATION STUDY OF THE ERROR INDUCED IN
ONE-SIDED RELIABILITY CONFIDENCE BOUNDS FOR
THE WEIBULL DISTRIBUTION USING A SMALL
SAMPLE SIZE WITH HEAVILY CENSORED DATA**

by

Michael A. Hartley

December 2004

Thesis Advisor:
Second Reader:

David H. Olwell
Lyn R. Whitaker

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

| | | | | |
|---|---|--|--|--|
| REPORT DOCUMENTATION PAGE | | | <i>Form Approved OMB No. 0704-0188</i> | |
| Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503. | | | | |
| 1. AGENCY USE ONLY (Leave blank) | | 2. REPORT DATE December 2004 | 3. REPORT TYPE AND DATES COVERED Master's Thesis | |
| 4. TITLE AND SUBTITLE: A Simulation Study of the Error Induced in One-sided Reliability Confidence Bounds for the Weibull Distribution Using a Small Sample Size with Heavily Censored Data | | | 5. FUNDING NUMBERS | |
| 6. AUTHOR(S) Michael A. Hartley | | | | |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000 | | | 8. PERFORMING ORGANIZATION REPORT NUMBER | |
| 9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A | | | 10. SPONSORING/MONITORING AGENCY REPORT NUMBER | |
| 11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. | | | | |
| 12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited | | | 12b. DISTRIBUTION CODE A | |
| <p>Budget limitations have reduced the number of military components available for testing, and time constraints have reduced the amount of time available for actual testing resulting in many items still operating at the end of test cycles. These two factors produce small test populations (small sample size) with "heavily" censored data. The assumption of "normal approximation" for estimates based on these small sample sizes reduces the accuracy of confidence bounds of the probability plots and the associated quantities. This creates a problem in acquisition analysis because the confidence in the probability estimates influences the number of spare parts required to support a mission or deployment or determines the length of warranty ensuring proper operation of systems. This thesis develops a method that simulates small samples with censored data and examines the error of the Fisher-Matrix (FM) and the Likelihood Ratio Bounds (LRB) confidence methods of two test populations (size 10 and 20) with three, five, seven and nine observed failures for the Weibull distribution. This thesis includes a Monte Carlo simulation code written in S-Plus that can be modified by the user to meet their particular needs for any sampling and censoring scheme. To illustrate the approach, the thesis includes a catalog of corrected confidence bounds for the Weibull distribution, which can be used by acquisition analysts to adjust their confidence bounds and obtain a more accurate representation for warranty and reliability work.</p> | | | | |
| 14. SUBJECT TERMS: Weibull Distribution, Simulation Study, Confidence Interval, Small Sample Size, S-Plus, Monte Carlo Simulation. | | | NUMBER OF PAGES 79 | |
| | | | 16. PRICE CODE | |
| 17. SECURITY CLASSIFICATION OF REPORT Unclassified | 18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified | 19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified | 20. LIMITATION OF ABSTRACT UL | |

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited.

**A SIMULATION STUDY OF THE ERROR INDUCED IN ONE-SIDED
RELIABILITY CONFIDENCE BOUNDS FOR THE WEIBULL DISTRIBUTION
USING A SMALL SAMPLE SIZE WITH HEAVILY CENSORED DATA**

Michael A. Hartley
Civilian, Department of the Air Force
MSEE, Air Force Institute of Technology, 1988

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN APPLIED SCIENCE

from the

**NAVAL POSTGRADUATE SCHOOL
December 2004**

Author: Michael A. Hartley

Approved by: Professor David H. Olwell
Thesis Advisor

Professor Lyn R. Whitaker
Second Reader

Professor James N. Eagle
Chairman, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Budget limitations have reduced the number of military components available for testing, and time constraints have reduced the amount of time available for actual testing resulting in many items still operating at the end of test cycles. These two factors produce small test populations (small sample size) with heavily censored data. The assumption of normality for estimates based on these small sample sizes reduces the accuracy of confidence bounds of the probability plots and the associated quantities. This creates a problem in acquisition analysis because the confidence in the probability estimates influences the number of spare parts required to support a mission or deployment or determines the length of warranty ensuring proper operation of systems. This thesis develops a method that simulates small samples with censored data and examines the error of the Fisher-Matrix (FM) and the Likelihood Ratio Bounds (LRB) confidence methods of two test populations (size 10 and 20) with three, five, seven and nine observed failures for the Weibull distribution. This thesis includes a Monte Carlo simulation code written in S-Plus that can be modified by the user to meet their particular needs for any sampling and censoring scheme. To illustrate the approach, the thesis includes a catalog of corrected confidence bounds for the Weibull distribution, which can be used by acquisition analysts to adjust their confidence bounds and obtain a more accurate representation for warranty and reliability work.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

| | | |
|-------------|---|-----------|
| I. | INTRODUCTION..... | 1 |
| A. | OVERVIEW | 1 |
| B. | PURPOSE..... | 2 |
| C. | SCOPE AND LIMITATIONS..... | 3 |
| II. | BACKGROUND | 5 |
| A. | LIFE DATA ANALYSIS..... | 5 |
| 1. | Confidence Bounds Defined | 5 |
| 2. | One-Sided Confidence Bounds | 6 |
| 3. | Non-Symmetric Life Distributions in Small Samples..... | 6 |
| 4. | Censoring | 7 |
| 5. | The Weibull Distribution | 7 |
| 6. | Fisher-Matrix Method | 9 |
| 7. | One-Sided Confidence Bounds for the Weibull Distribution Using the Likelihood Ratio Method | 10 |
| III. | METHODOLOGY | 13 |
| A. | SIMULATION | 13 |
| B. | APPLICATION..... | 15 |
| IV. | RESULTS | 19 |
| A. | INTRODUCTION..... | 19 |
| B. | SUMMARY OF GRAPHS..... | 19 |
| C. | ERROR CORRECTION USING THE CATALOG | 22 |
| D. | CODE MODIFICATION FOR SPECIFIC APPLICATIONS | 24 |
| E. | TEST CASE..... | 24 |
| V. | CONCLUSIONS AND RECOMMENDATIONS..... | 29 |
| A. | CONCLUSIONS | 29 |
| B. | RECOMMENDATIONS..... | 30 |
| | APPENDIX A | 31 |
| | APPENDIX B | 33 |
| A. | RELIABILITY = 99%, SAMPLE SIZE = 10 | 33 |
| B. | RELIABILITY = 99%, SAMPLE SIZE = 20 | 35 |
| C. | RELIABILITY = 98%, SAMPLE SIZE = 10 | 37 |
| D. | RELIABILITY = 98%, SAMPLE SIZE = 20 | 39 |
| E. | RELIABILITY = 97%, SAMPLE SIZE = 10 | 41 |
| F. | RELIABILITY = 97%, SAMPLE SIZE = 20 | 43 |
| G. | RELIABILITY = 96%, SAMPLE SIZE = 10 | 45 |
| H. | RELIABILITY = 96%, SAMPLE SIZE = 20 | 47 |
| I. | RELIABILITY = 95%, SAMPLE SIZE = 10 | 49 |
| J. | RELIABILITY = 95%, SAMPLE SIZE = 20 | 51 |
| K. | RELIABILITY = 90%, SAMPLE SIZE = 10 | 53 |

| | |
|--|----|
| L. RELIABILITY = 90%, SAMPLE SIZE = 20 | 55 |
| LIST OF REFERENCES..... | 57 |
| INITIAL DISTRIBUTION LIST | 59 |

LIST OF FIGURES

| | | |
|------------|--|----|
| Figure 1. | Two-sided 95% confidence interval symmetrically distributed around the mean for a normal distribution..... | 6 |
| Figure 2. | One-sided upper 95% confidence interval for the mean for a normal distribution. | 7 |
| Figure 3. | Weibull probability density functions for different shape parameters with a fixed scale parameter. | 8 |
| Figure 4. | Weibull hazard functions showing the effect of changing the shape parameter with a fixed scale parameter. | 9 |
| Figure 5. | 2-Parameter Weibull likelihood distribution showing a 95% confidence level (horizontal line) and confidence bounds (vertical lines) of (1.897, 4.772) [Ref: 3, p. 181]..... | 11 |
| Figure 6. | A 95% confidence interval for normally distributed data with a sample size of 10, showing one of the 40 CI's doesn't contain the true value for an observed coverage of 95%. [Ref: 9] | 14 |
| Figure 7. | A 90% confidence interval for normally distributed data with a sample size of 10, showing four of the 40 intervals do not contain the true value for an observed coverage of 90%. [Ref: 9] | 14 |
| Figure 8. | A 95% confidence interval for normal distributed data with a sample size of 20, showing that two of the 40 intervals do not contain the true value for an observed coverage of 95%. [Ref: 9] | 15 |
| Figure 9. | Example graph generated in S-Plus using a 1200 iteration Monte Carlo simulation showing the simultaneous plots of the FM and LRB methods. | 16 |
| Figure 10. | Expanded view of a typical graph showing the intersection of the actual coverage with each method and the resulting requested coverage. | 17 |
| Figure 11. | Example graph generated in S-Plus using a modified code that displays the required values and automatically draw the lines. | 18 |
| Figure 12. | A graph showing an increasing error in the confidence bounds as the numbers of failures decrease for the Fisher-Matrix Method for 98% reliability and a test population of 10. | 20 |
| Figure 13. | Output of the S-Plus simulation code using N =13 and f =10 derived from Example 4.1.1. [Ref: 10]..... | 26 |
| Figure 14. | The output of the Weibull++6 software package used to determine the B10C95 value for the S-Plus simulation code. | 27 |

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

| | | |
|----------|---|----|
| Table 1. | Percent error in the confidence bounds using the Fisher-Matrix Method (FM), a sample size of 10, and an 80% (actual) confidence level. (Where the Error = $\text{abs}[(\text{Required}-\text{Actual})/\text{Actual}] \times 100\%$). | 21 |
| Table 2. | Percent error in the confidence bounds using the Likelihood Ratio Bounds (LRB) method, a sample size of 10, and an 80% (actual) confidence level. (Where the Error = $\text{abs}[(\text{Required}-\text{Actual})/\text{Actual}] \times 100\%$). | 22 |
| Table 3. | Summary of the required confidence using the Fisher-Matrix method (FM) and an actual 80% confidence using a Monte Carlo simulation with a sample size of 10 with 3, 5, 7, & 9 failures for 99%, 98%, 97%, 96%, 95%, and 90% reliability..... | 23 |
| Table 4. | Summary of the required confidence using the Fisher-Matrix method (FM) and an actual 80% confidence using a Monte Carlo simulation with a sample size of 20 with 3, 5, 7, & 9 failures for 99%, 98%, 97%, 96%, 95%, and 90% reliability..... | 23 |
| Table 5. | Aircraft component failure times in hours. [Ref:10, p. 155] | 25 |
| Table 6. | Output from the Weibull++6 Life Data Analysis Software package used for comparison with Example 4.1.1. [Ref: 10, p. 155] | 28 |
| Table 7. | Output from the Weibull++6 Life Data Analysis Software package using the confidence value from the S-Plus simulation code for Example 4.1.1. [Ref: 10, p. 155]..... | 28 |

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First and foremost, I thank my wife Rhonda for her support in this endeavor. If I had your drive and your skill at organizing, categorizing, and sorting, I would have finished before I began!

My sincerest thanks go to my advisor, Prof. David Olwell, for the initial concept for this thesis and the continuing support throughout the writing. This topic is of great interest to him and his enthusiasm was contagious as my interest grew and the application to relevant work issues expanded. I appreciate both the technical direction and the encouragement he gave in cajoling me to “Write like the wind!”

In addition, I would like to thank Prof. Lyn Whitaker for taking time from her teaching schedule to review and comment on this paper. Your first course in Statistics peaked my interest in this subject and allowed me the ‘confidence’ to finish the thesis.

Finally, I would like to thank Prof. Sam Buttrey (“Sam I am”) who developed the S-Plus code for the simulations. It was a pleasure to watch you ‘drive’ as the code magically appeared on the screen. You-da-man!

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

Shrinking budgets have affected all aspects of military hardware testing including the number of items that are procured for test in order to establish and verify their reliability. The amount of time given to test specific items has been reduced to shorten the acquisition time. These two factors contribute to fewer numbers of test items put into a test scenario with even a smaller number of failures occurring during the allotted test time.

The few failures that occur during the test period provide valuable information to analysts, but the ones that don't fail (the survivors at the end of the allotted test time) may provide even more information. The items that survive the test period are referred to as "censored" and in particular, the items whose testing begins at the same time and ends when the allotted test time is completed is referred to as "time-censored" or Type I censored data. Another common type of censored data is "failure censored" or Type II and may be thought of as ending the test when a predetermined number of failures have occurred. Estimates based on Type II censoring are often used to generate estimates under Type I censoring because the tester can control the number of failures and hence, the number of censored items.

Many current software packages such as Weibull++, S-Plus, and Minitab allow the analyst the ability to specify the total number of test items put into test, the times of the failures for any items, and the number of censored items (items still running at the end of the test time). The software packages listed above are based on large sample theory, yet users readily apply the analysis to a small sample size with heavily censored data with little regard to the inherent error.

This assumption of normality due to large sample size, when applied to testing based on small sample sizes, has been known for decades to provide errors in the system reliability estimates and the confidence associated with those estimates. These errors were tolerated as minor, but in today's economy of doing more with less, the government must be a good steward of its money and demand more accurate results from the limited assets it's given.

The confidence associated with a system reliability estimate influences the number of spare parts required to support a mission or deployment or determines the length of warranty ensuring proper operation of systems. To ensure accurate confidence in small sample testing, this thesis develops a Monte Carlo methodology for the Weibull distribution, simulating small samples with censored data and examines the error of two commonly used methods to determine confidence bounds for reliability: 1. the Fisher-Matrix (FM) method and 2. the Likelihood Ratio Bound (LRB) method. Tables are provided that summarize the graphical output of the simulation code and quantify the difference between the two methods, with the end result of providing the user a more accurate method.

A simulation method is used in this thesis to quantify the bias in confidence bounds for reliability that one obtains when using software packages based on large sample theory and applies that analysis to small sample size tests. A simulation of two small-sample test populations (sample size of 10 and 20) is used with three, five, seven and nine observed failures for the Weibull distribution. Type II censoring is used during the simulation by specifying the number of failures that occur out of the total sample size. The S-Plus code by default performs a 10000 iteration, Monte Carlo simulation for each test case and produces a graph for the more often used Fisher-Matrix (FM) confidence method. A catalog (Appendix B) is developed that can provide the user a more accurate confidence that they can use in their acquisition documents for many common reliabilities and censored data combinations.

A detailed process is presented that leads a user through an example clarifying the simulation process. This example provides an avenue for the user to take advantage of the attached catalog showing several common reliability levels and confidence bounds for the Weibull distribution, and allows them to literally ‘pick-off’ the adjusted confidence bounds for their particular situation. In addition, the documented S-Plus code is included that can be modified to meet the users requirements allowing the graphs to be tailored to specific needs.

The errors produced by the naïve use of large sample methods are substantial when applied to small samples. A requested 90% lower confidence bound on reliability in

typical software packages may only provide an actual 75% coverage for small samples. This thesis corrects that error by allowing the user to request a corrected nominal confidence level for small sample testing to achieve the true desired coverage when using software packages that are based on the “large sample theory”. For example, we may need to request a 99.2% confidence level to actually get a 90% coverage, for a given sample size and censoring scheme.

This ability to correct for the error in the confidence levels allows us to continue to use existing software tools, but to use them better.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW

Confidence intervals (CI) in life data analysis are an indication of the degree of confidence that the analyst presents along with the statistical data. It is his “confidence” that a specific interval contains the quantity he is interested in. As stated quite nicely in their text on reliability methods, Meeker and Escobar said [Ref: 3, p.49],

A specific interval either contains the quantity of interest or not; the truth is unknown.

Therefore, the confidence interval **quantifies** the uncertainty, but as Nelson states in his book on applied life data analysis [Ref: 4, p.196],

The real uncertainty is usually greater than the confidence interval indicates because the interval is based on certain assumptions about the data and the model...Departures from the assumptions add to the uncertainties in the results.

Most statistical analysis (and life data analysis) assumes asymptotic normality of estimators based on large sample size theory, but as seen in recent acquisition testing, the government rarely has large numbers of test articles. Small sample sizes are especially prevalent in Developmental Testing (DT) and Operational Testing (OT) in procuring, testing, and buying major weapon systems. In addition to a small number of test articles that are provided for testing, the contractors generally have good in-house reliability programs that result in few failures during the test period; this leads to censored data.

Even if a sufficient number of test items are available to meet the criteria of the large-sample size theory, the number of failures need to be large [Ref: 3, p.186], not just the number of test items. As Nelson points out in his book on accelerated testing [Ref: 1, p.236],

Such asymptotic theory is also called **large-sample** theory. This terminology is misleading for multiply and singly censored data, since the number of *failures* needs to be large. In practice, the asymptotic theory is applied to small samples, since crude theory is better than no theory.

This error in confidence bounds for system reliability using small sample sizes has been known and accepted for years. The impact of these errors in today’s climate of ‘doing more with

less' and 'the sooner the better' impacts all aspects of life cycle cost estimation and possibly even mission capability. Budget limitations and time constraints force the tester to use a smaller number of test items with less time to test them, which in turn results in a small sample size that is heavily censored. The analyst must then apply statistical methods that are inherently biased for small samples. In his text on probability and statistics, Devore reflects on his concerns about using asymptotic methods on small samples when the distribution is non-normal [Ref: 5, p.302],

It would certainly be distressing to believe that your confidence level is about 95% when in fact it is really more like 88%!

The government as a consumer should get the confidence it pays for and this thesis allows for the correct adjustment to the contractual documents before products are purchased. Most commercial software packages (Weibull ++, Minitab, WinSmith, S-Plus) available for life data analysis use a normal-approximation method to determine the confidence intervals even for small sample sizes because the large-sample approximations are computationally easy. [Ref: 3, p. 186, Ref: 6, p. 135].

Many researchers have documented that a bias exists with small samples and have focused on finding new methods to improve the coverage of confidence intervals for reliability estimates(e.g. Nelson (1982), Piergorsch (1987), Ostrouchov and Meeker (1988), Vander Weil and Meeker (1990), Doganaksoy and Schmee (1993), Shao and Tu (1995), Jeng and Meeker (2000) [Ref: 6, p.135-136]).

This thesis suggests using existing methods better by correcting for the bias in existing commercial software products.

B. PURPOSE

The purpose of this thesis is to document the error in confidence interval coverage for reliability estimates based on data from a Weibull distribution between the Fisher-Matrix (FM) and Likelihood Ratio Bounds (LRB) methods by conducting a simulation study and constructing graphs and tables that indicate the extent of the problem. A 10000 iteration, Monte Carlo simulation will be used to produce graphs that analysts can use to convert the nominal $(1-\alpha)$ 100% 1-sided confidence bounds into confidence bounds with actual $(1-\alpha)$ 100% coverage. In

addition, the S-Plus code using the embedded SPLIDA routine [Ref: 2] that was written for the simulation study is included in Appendix A and allows users to tailor a simulation to fit their specific needs.

Once the error is known, it can be corrected by applying an appropriate adjustment, resulting in the desired coverage level. This correction is determined both graphically and numerically.

C. SCOPE AND LIMITATIONS

Most texts on life data analysis related to reliability and warranty calculations suggest that using the large sample approximations to get approximate confidence intervals is good for quick and easy calculations [Ref: 3, p.165], but more accurate procedures are required for formal proposals and final reports. This thesis performs a Monte Carlo simulation study using 10000 iterations determining actual ‘coverage probabilities’ of confidence intervals for system reliability based on small samples and heavily censored data. These actual coverage probabilities can be used to correct the error in the asymptotic methods, providing the accurate procedures desired for formal proposals and final reports.

This simulation method works for any location-scale or log-location-scale distribution. This thesis uses Weibull for simplicity, but the simulation code can be adapted for lognormal, normal, smallest extreme value, exponential, and logistic distributions.

The thesis focuses on using a simulation code to quantify the error between the two confidence methodologies that were developed for large samples but are applied to small samples and will be limited to calculating the exact lower, one-sided confidence interval for the following test cases:

Typical reliability values of 99, 98, 97, 96, 95, and 90 will be used in the simulation study.

Small sample sizes of 10 and 20 with 3, 5, 7, and 9 observed failures will be used for each reliability value.

The Weibull location-scale distribution with fixed parameters will be used exclusively for this simulation study.

In addition, the Fisher-Matrix (FM) bounds and Likelihood Ratio Bounds (LRB) methods will be compared to obtain the bias in confidence interval coverage. Results of the comparison will be documented in tables and will establish which method has the most error. A catalog of the method with the most error will be attached as Appendix B for user convenience.

And finally, the S-Plus code for the Monte Carlo simulation will be attached as Appendix A allowing users to tailor the simulation output for their specific application. The attached S-Plus code also provides the user a subroutine that states the exact confidence level they must request requested in order to obtain the desired coverage, for a given sampling size, censoring scheme, and distribution. This eliminates the need to estimate the required value using the graphs in Appendix B.

II. BACKGROUND

A. LIFE DATA ANALYSIS

Life data analysis seeks to understand the data at hand obtained from a sample of test articles (descriptive statistics) and then apply the results to the entire population of test articles (inferential statistics). This section is provided as background material to focus the reader on the specific terms, procedures and definitions used in this thesis and provide a common background for users who want to tailor the S-Plus code for their specific application.

1. Confidence Bounds Defined

To compute a confidence interval from a given data set, you must first specify a confidence level. For example, if a data set consists of n observations assumed to have a normal distribution with known variance and unknown mean and if you want to find the 95% confidence interval for the mean, the equation for the confidence interval endpoints (**confidence bounds**) would look like (lower endpoint, upper endpoint) and is represented mathematically as:

$$\left(\bar{x} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} \right) \quad (2.1)$$

where σ is the known standard deviation, \bar{x} is the mean of the data, α is $(1 - .95)$ and n is the number of observations. The value of 1.96 (used for $z_{\alpha/2}$) is obtained from the standard normal probability tables and represents the critical value under the normal curve that leaves the center area equal to .95 (the level of confidence desired). The area under the tails is the $(1 - .95)*100\% = 5\%$ that is equally divided for symmetric distributions (shown in Figure 1). The equal distribution of the confidence on either end of the probability curve displayed in Figure 1 is only correct for a symmetric distribution such as the normal family.

According to Devore in his text on probability and statistics [Ref: 5, p. 281], the correct interpretation of this 95% confidence bound is true only over the long run,

To say that an event A has probability (of) .95 is to say that if the experiment on which A is defined is performed over and over again, in the long run A will occur 95% of the time.

In other words, the 95% confidence level is not so much a statement of fact about any specific interval, but represents what would happen if a large number of similar events took place. This statement leads to the Monte Carlo technique of using 10000 iterations in the simulation study used in this thesis.

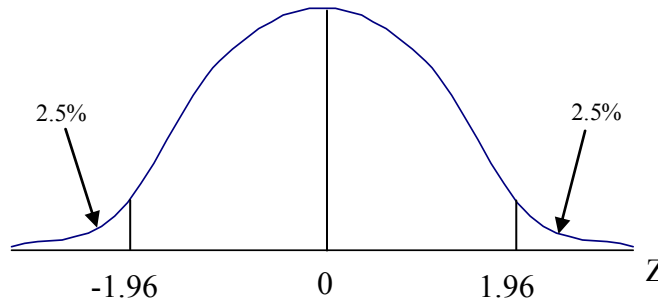


Figure 1. Two-sided 95% confidence interval symmetrically distributed around the mean for a normal distribution.

2. One-Sided Confidence Bounds

Most reliability and warranty analysts use a one-sided confidence interval for reliability [Ref: 5, p.292] because they are interested in estimating the reliability with some confidence that the actual value is at least as large as that lower bound. In terms of risk, a one-sided interval establishes, with some confidence, the largest (or smallest) plausible value. For a symmetric distribution like the normal curve, the adjustment between a two-sided and a one-sided CI is made by replacing $z_{\alpha/2}$ with z_{α} and choosing the appropriate end of the distribution (either the + or the - in Eq. 2.1). Now the single area to the right of the critical value 1.64 represents $(1 - .95) \cdot 100\% = 5\%$ as shown in Figure 2.

3. Non-Symmetric Life Distributions in Small Samples

When the sample size is small, the sampling distribution of an estimator may be skewed (a non-symmetric distribution) [Ref: 3, p.56] and using the normal approximation for confidence levels near either tail of the distribution may give a substantial error.

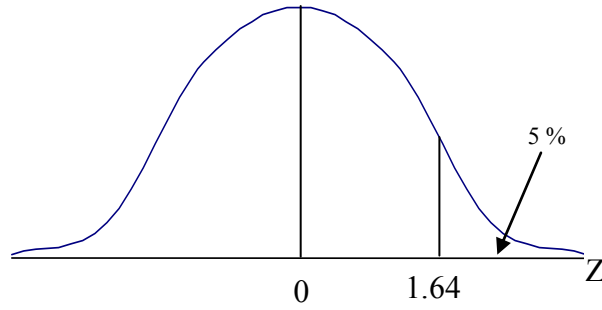


Figure 2. One-sided upper 95% confidence interval for the mean for a normal distribution.

4. Censoring

When several items are put into a test environment, either simultaneously or sequentially, not all of the units may fail by the end of the test. The survivors (unfailed units) are called censored, in that incomplete failure information is available; you only know they were still functioning at the end of the allotted test time. Type I censoring is generally considered as removing unfailed units from a test at a prescribed time (i.e. at the end of the allotted test time). Type I is also called “time censoring” by some authors and is considered to be the more common type of censoring in reliability data analysis. [Ref: 3, p.7; Ref: 6, p.1] Type II censoring is called “failure censoring” and is based on ending the test when a predetermined number of failures occur.

5. The Weibull Distribution

The Weibull distribution is a parametric distribution that is valuable in modeling certain life data. The benefit of using parametric models is that the distribution can be expressed entirely by just a few parameters. Caution should be used in referencing texts in probability and statistic theory that explain the Weibull distribution as the notation representing the shape and scale parameters are not standardized. This thesis uses the parameters β and η for the shape and scale respectively as is shown in Meeker and Escobar’s text on statistical methods for reliability data [Ref: 3, p. 85]. The following equation is the probability density function (pdf) for the Weibull distribution using β and η :

$$f(t; \beta, \eta) = \frac{\beta}{\eta^\beta} t^{\beta-1} \exp \left[- \left(\frac{t}{\eta} \right)^\beta \right], \quad t \geq 0, \text{ and } \beta, \eta > 0 \quad (2.2)$$

Integrating the pdf with respect to the independent variable (t) yields the easily recognized cumulative distribution function (cdf) for the Weibull distribution:

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right], \quad t \geq 0 \quad (2.3)$$

The system reliability $R(t)$ is the probability that an item survives past time t and is defined as $R(t) = 1 - F(t)$.

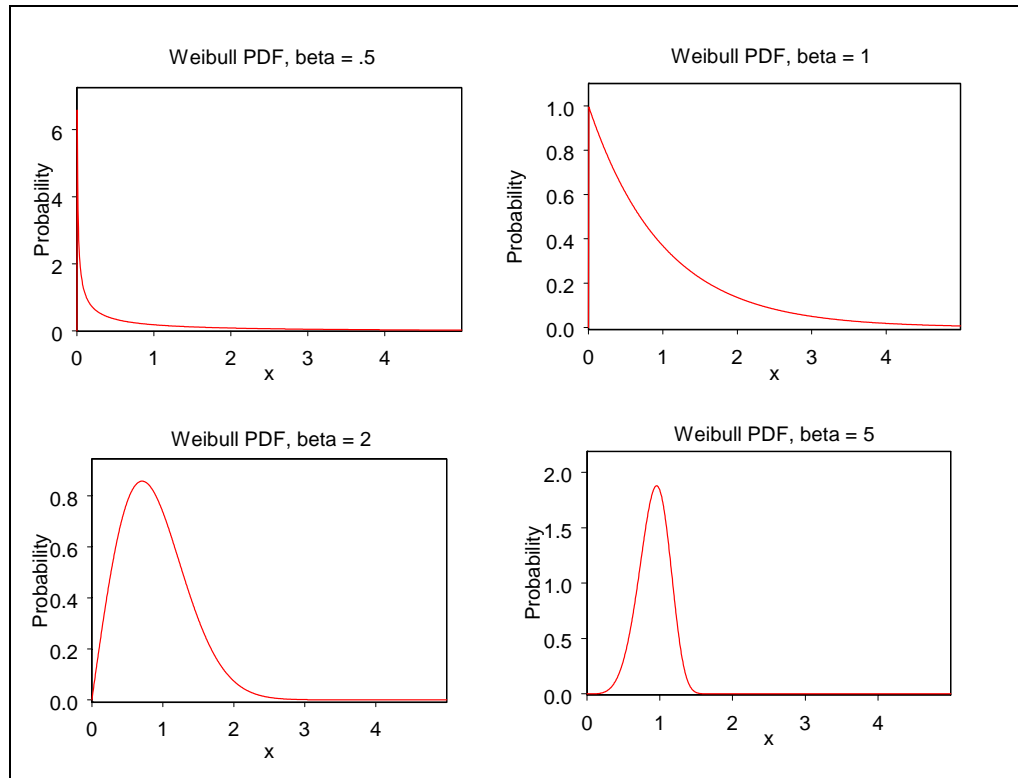


Figure 3. Weibull probability density functions for different shape parameters with a fixed scale parameter.

Figure 3 shows that changing the value of the shape parameter allows the Weibull distribution to model many different types of behavior. The Weibull distribution can have either a decreasing or an increasing hazard function depending on the choice of parameters making it the “workhorse” of life data analysis because of its wide application. [Ref: 3, p.173]

The hazard function is defined as the “propensity for an item to fail in the next small interval of time, given that the item has survived until some time (t)” [Ref: 3, p. 28] and is given by the equation $h(t) = f(t)/(1-F(t))$, where $f(t)$ is the probability density function (pdf) and $F(t)$ is the cumulative distribution function (cdf). The following set of plots in Figure 4 use the same shape parameter values as the Weibull probability density functions above, but displays the hazard function.

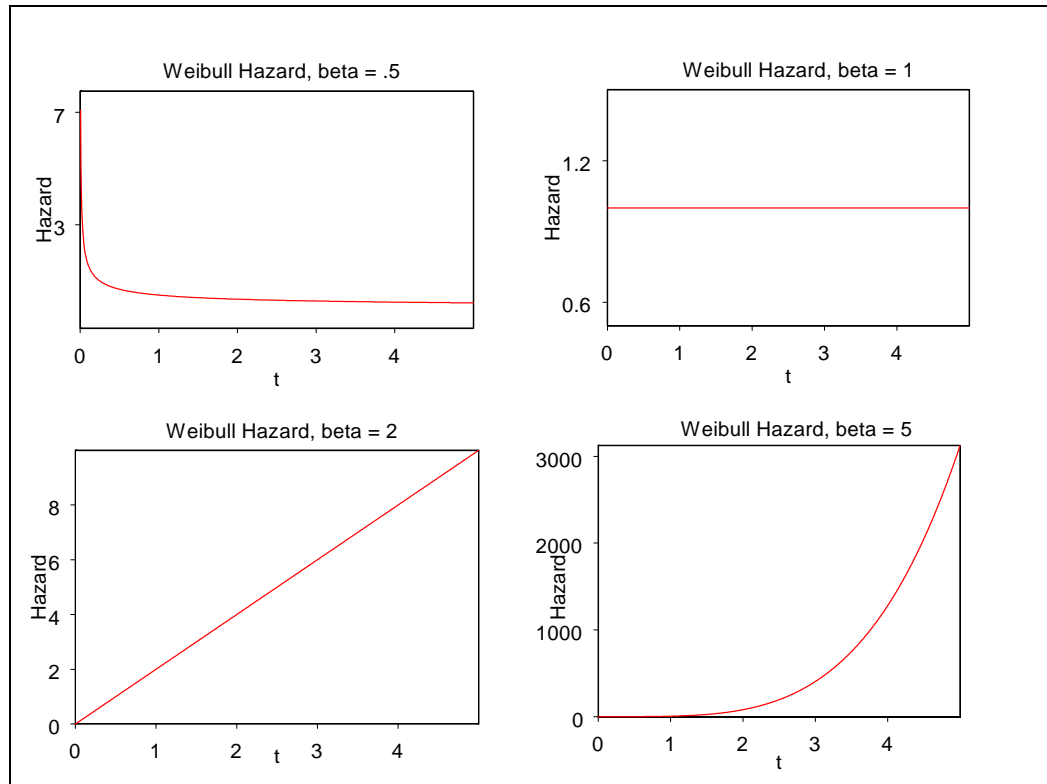


Figure 4. Weibull hazard functions showing the effect of changing the shape parameter with a fixed scale parameter.

6. Fisher-Matrix Method

Confidence intervals (and hence the confidence bounds) can be calculated for parameters of the Weibull distribution in a manner similar to that shown for the normal distribution in Equation 2.1. Typical software packages for life data analysis use the normal-approximation method because it's fast and easy. The best practice for finding the confidence interval for the shape parameter β is to perform a logarithmic transformation on the estimator $\hat{\beta}$ to improve the

symmetry of its sampling distribution, and then approximate the sampling distribution of $\ln(\hat{\beta})$ by a normal distribution. The confidence bounds then become,

$$[lowerbound(L), upperbound(U)] = [\hat{\beta} / W, \hat{\beta} * W] \quad (2.4)$$

where $W = \exp[z_{(1-\alpha/2)} \hat{\sigma}_{\hat{\beta}} / \hat{\beta}]$ and $\hat{\sigma}_{\hat{\beta}} = \sqrt{\hat{Var}(\hat{\beta})}$. The value for $\hat{Var}(\hat{\beta})$ is obtained from the parameter variance-covariance matrix that is calculated in the analysis software using the standard maximum likelihood estimates (MLE) [Ref: 3, Appendix B]. For a symmetric distribution such as the normal, the adjustment between a two-sided and a one-sided CI is made by replacing $z_{\alpha/2}$ with z_{α} and choosing the appropriate end of the distribution (either the upper bound or the lower bound in Eq. 2.4). Similar adjustments are made to compute confidence intervals for other parameters and functions of parameters, such as reliability.

7. One-Sided Confidence Bounds for the Weibull Distribution Using the Likelihood Ratio Method

Confidence intervals based on a symmetric sampling distribution produce equal values in the tails of the probability density functions (pdf). However, often the sampling distribution for an estimator (such as $\hat{\beta}$) is not symmetric. The usual sampling distribution in most practical applications takes a shape similar to the southeast corner example in Figure 3. To find the approximate two-sided confidence interval, the relative likelihood is used. The justification for this approach is presented in various texts [Ref: 3, p. 627] and leads to the following equation for determining a two-sided confidence interval:

$$R(\theta) \geq \exp[-\chi^2_{(1-\alpha;1)} / 2] \quad (2.5)$$

where $R(\theta)$ is the ratio of the likelihood function of a single parameter (theta) divided by the likelihood evaluated at the maximum likelihood value of the parameter. Values of the Chi-Squared distribution are available in tables in most statistic texts.

The technique for finding the two-sided interval can be envisioned as plotting the relative likelihood (on the left y-axis) against the values of the estimated parameter while simultaneously plotting the confidence levels (on the right y-axis). The value of Eq. 2.5 determines a horizontal line that intersects the graph of the relative likelihood function (see Figure 5). The vertical lines

drawn from the intersections of the horizontal line and the relative likelihood, down to the x-axis, determine the confidence interval endpoints.

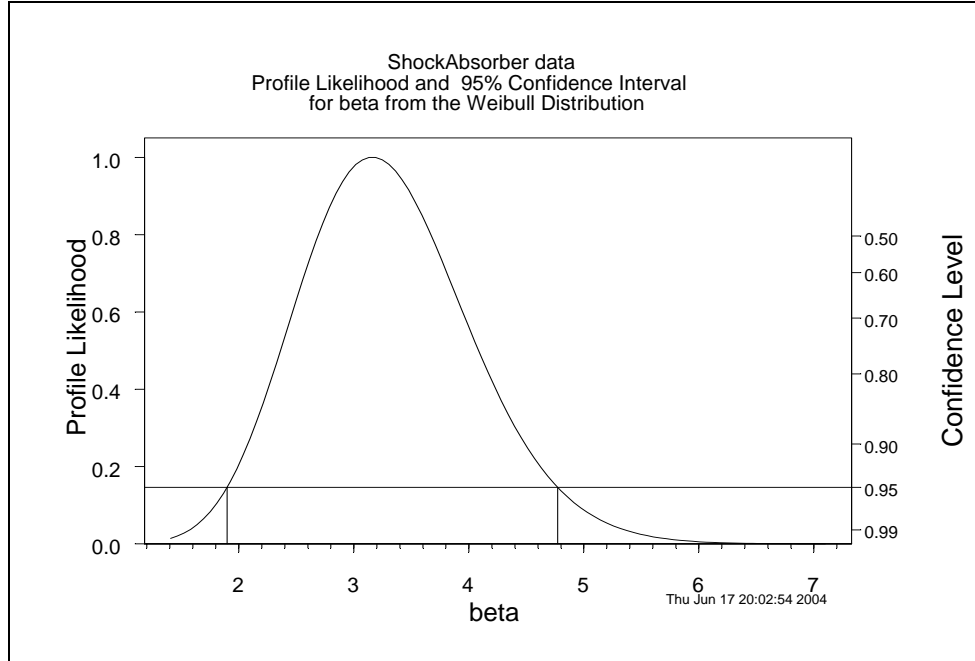


Figure 5. 2-Parameter Weibull likelihood distribution showing a 95% confidence level (horizontal line) and confidence bounds (vertical lines) of (1.897, 4.772) [Ref: 3, p. 181]

(Note that the area is not symmetrically distributed between the upper and lower bounds.)

This thesis is interested in finding the one-sided confidence bound of for reliability based on a Weibull distribution as is often used in reliability measurement and warranty calculations. A one-sided confidence interval for θ can be obtained from Eq. 2.5 by replacing α with 2α and drawing the horizontal line in Figure 5 at the value where

$$R(\theta) \geq \exp\left[-\chi^2_{(1-2\alpha;1)} / 2\right] \quad (2.6)$$

Then, choosing the correct endpoint from the x-axis of relative likelihood graph, you can obtain either the lower or the upper, one-sided confidence bound. This is the usual practice for one-sided likelihood ratio confidence bounds. [Ref: 3, p. 186]

This approach extends to distributions that have more than one parameter and to functions of parameters, and is discussed extensively by Meeker and Escobar [Ref: 3, Chapter 8]. The key point is that the relative likelihood is asymmetric, and so two-sided intervals using it have better coverage properties than the FM method. For one-sided confidence intervals, use of Equation 2.6 still presents problems as it assumes that the confidence is distributed evenly between the upper and lower tails of the region. This is not the case, and is the source of the problems for this method when applied to one-sided intervals.

III. METHODOLOGY

A. SIMULATION

This thesis uses a reliability software package called S-Plus [Ref: 7] to simulate small sample size experiments with few failures. S-Plus uses an attached life data analysis routine called SPLIDA [Ref: 2] that performs the life data analysis. The S-Plus software package was chosen because of its wide application in the study of probability and statistics, its built-in life data analysis package, and the fact that S-Plus has the capability of user defined ‘functions’ that can be run for a large number of trials, hence the basis for the simulation study.

A simulation code (Appendix A) was written in S-Plus for the Weibull distribution that takes the location-scale parameters, beta (β) and eta (η), the number of items under test (n), an alpha (α) where $(1 - \alpha) * 100\%$ is the reliability required for a product (target or contractual), and the number of failures (f) recorded during the test time. The code first derives the true value of the quantile for the Weibull distribution using the given parameters and uses that value for comparing whether the result of each iteration yields a CI that contains that true value. If the true value is contained within the interval, a counter is incremented by one; if it is not contained within the interval it is not incremented. The ratio of the count to the number of iterations is an unbiased estimate of the probability coverage [Ref: 3, p. 49] and is plotted for each reliability value. An example of probability coverage is shown in the figures below.

Figure 6 is a graph of 40 simulations of a sampling from a normal distribution [Ref: 9] with a two-sided 95% CI for the mean calculated for each simulation. Compare that to Figure 7, which shows the same simulation run with a 90% CI. The width of each interval with 90% coverage is smaller than that of the 95% CI. Finally as a comparison, Figure 8 shows a 95% CI but with a larger sample size; it shows a small interval as well. The probability coverage represents a useful tool to determine if the value of interest falls within the CI and to quantify the coverage.

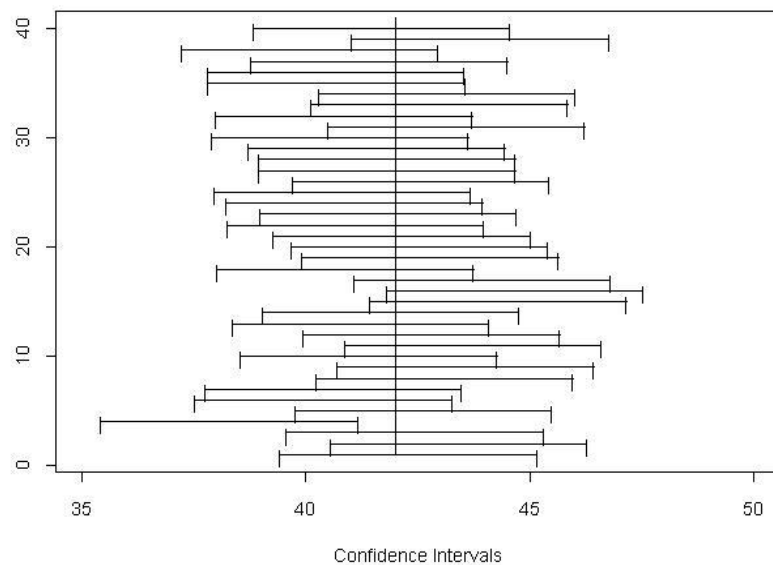


Figure 6. A 95% confidence interval for normally distributed data with a sample size of 10, showing one of the 40 CI's doesn't contain the true value for an observed coverage of 95%. [Ref: 9]

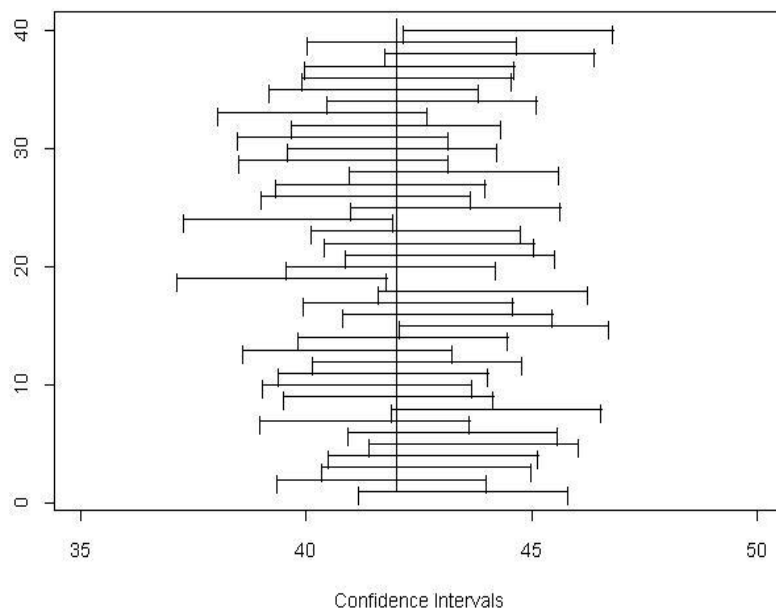


Figure 7. A 90% confidence interval for normally distributed data with a sample size of 10, showing four of the 40 intervals do not contain the true value for an observed coverage of 90%. [Ref: 9]

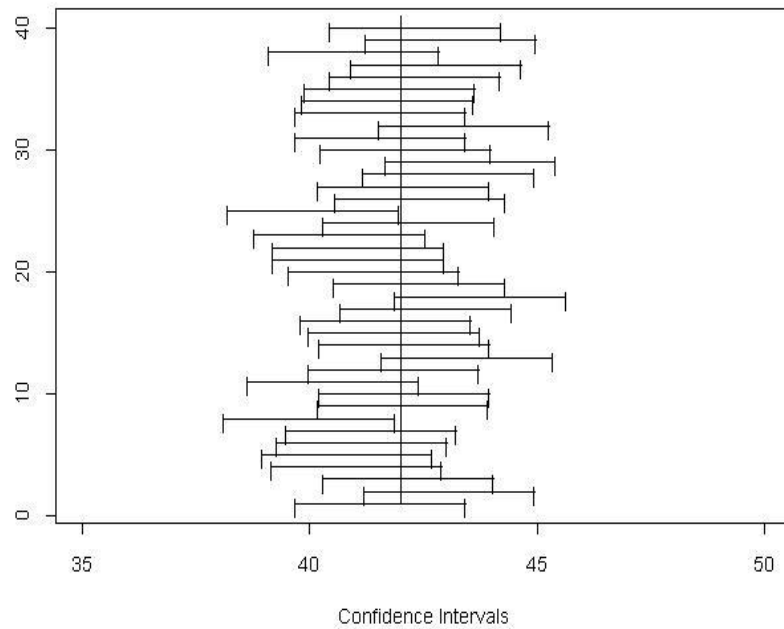


Figure 8. A 95% confidence interval for normal distributed data with a sample size of 20, showing that two of the 40 intervals do not contain the true value for an observed coverage of 95%. [Ref: 9]

The initial S-Plus code determined the probability coverage using both the Fisher-Matrix (FM) and the Likelihood Ratio Bound (LRB) methods and compared the two methods by simultaneously plotting them on the same graph. Even though both procedures are known to be inaccurate for small sample sizes with censored data, Meeker states [Ref: 3, p.165],

The computationally demanding likelihood procedure can be expected to provide better intervals (i.e. an actual coverage probability closer to nominal confidence level).

B. APPLICATION

The purpose of the simulation study was to quantify the bias between the two confidence methods and to develop tables showing the error. Then, the method with the greatest error was chosen to produce a catalog of graphs representing typical reliabilities often found in military contracts. The reliability is a required input to the simulation code that produces a separate graph for each value. The reliabilities chosen for this thesis were 99, 98, 97, 96, 95 and 90%. The numbers of items available for test (sample size of 10 and 20) establish the test population

and the numbers of failures that occur within each of the test populations (3, 5, 7, and 9) establish the amount of censoring. The confidence levels are not a required input to the simulations but are produced as a result of the S-Plus code for the two standard methods listed.

The output of the simulation produces a graph for each of the input variables (six reliabilities (times) two populations (times) four levels of censoring) resulting in a catalog of 48 graphs representing typical values found in most reliability and warranty specifications. Some graphs in this thesis display both of the two confidence methods simultaneously on each graph (see Figure 9 below for typical graph display) and allows a comparison between the two methods and is used to graphically display the error and differences between the two methods.

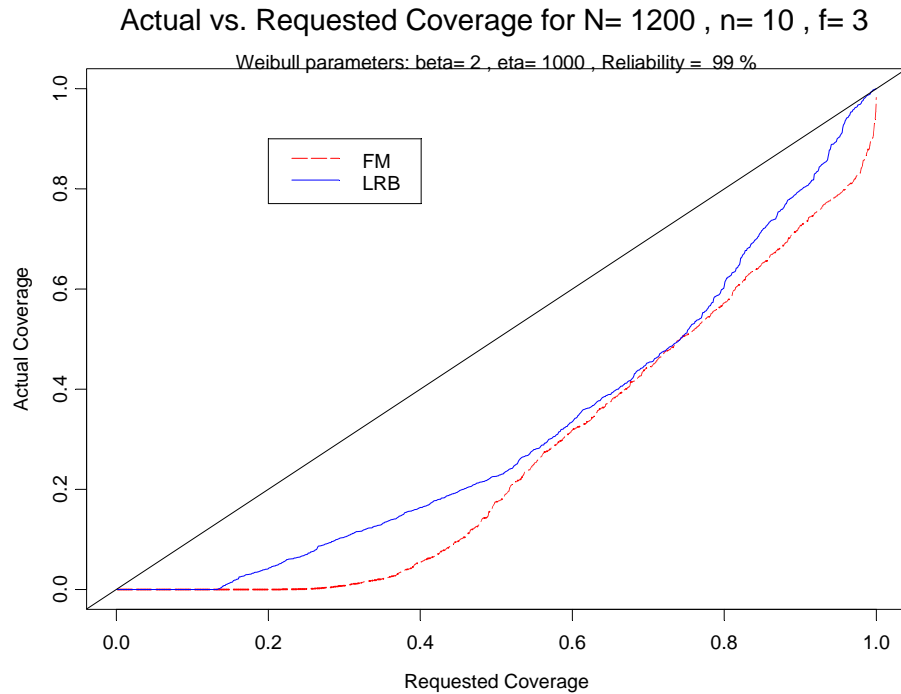


Figure 9. Example graph generated in S-Plus using a 1200 iteration Monte Carlo simulation showing the simultaneous plots of the FM and LRB methods.

The dashed line (red) on each graph is the simulation output of the one-sided confidence bound using the FM method and the blue line (solid) is the output using the LRB method. Most of the available software packages for life data analysis use the normal-approximation (Fisher-

Matrix) method with appropriate transformations to achieve symmetry but Meeker and Escobar assert that the likelihood ratio method is more accurate [Ref: 3, p. 165]. This thesis supports their assertion.

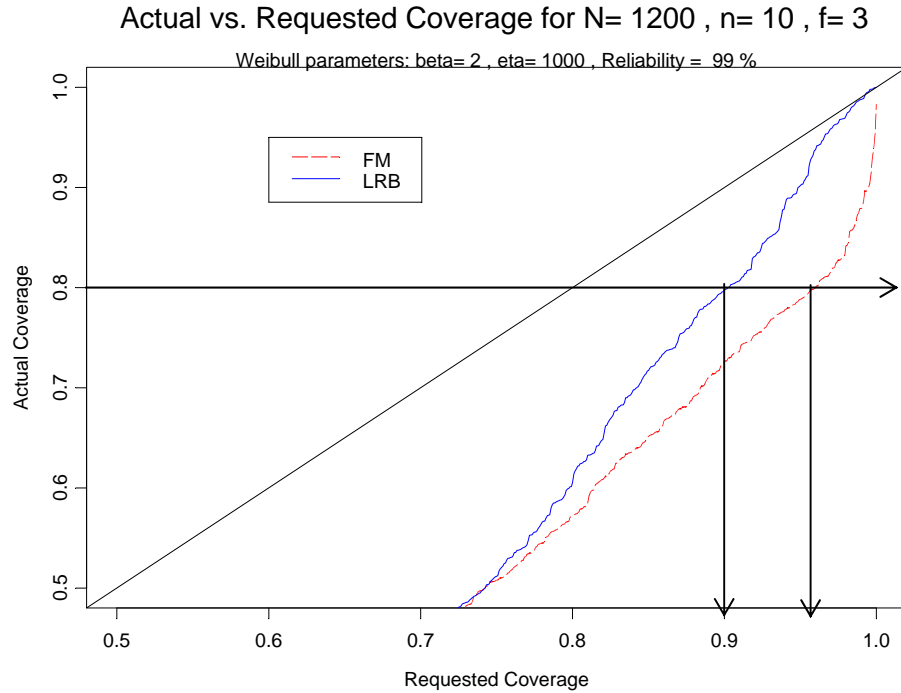


Figure 10. Expanded view of a typical graph showing the intersection of the actual coverage with each method and the resulting requested coverage.

The catalog of graphs is contained in Appendix B and only shows the more inaccurate Fisher-Matrix (FM) method. It is divided into six sections; one section for each reliability value with each section sub-divided by the sample size with a graph for each censoring level. The user need only turn to the appendix and select the correct reliability, find the sample size of interest, and locate the graph with the appropriate number of failures (censoring). The user then draws a horizontal line from the y-axis (Actual Coverage) to either one of the two confidence methods (solid or dashed lines), then down to the x-axis (Requested Coverage) to read the amount of confidence they require to have in order to get what he requests (see Figure 10). The black diagonal line across the middle of each graph is the theoretical value where the “Requested Confidence” equals the “Actual Confidence”. In other words, if we had a large enough samples and no censored data, both of the confidence methods would lie close to that line.

Figure 10 shows that as a result of running the simulation, that if the user needed an actual 80% one-sided confidence bound, they would have to request a 92% confidence bound using the LRB method and would need to request a 96% confidence bound using the FM method for this sample size and censoring scheme.

Figure 11 incorporates a modified S-Plus code that draws the line on the graph and displays the adjusted confidence bounds below the x-axis. The graph indicates that for an actual 90% coverage, you must request 92.16% for the LRB method and 96.478% for the FM method.

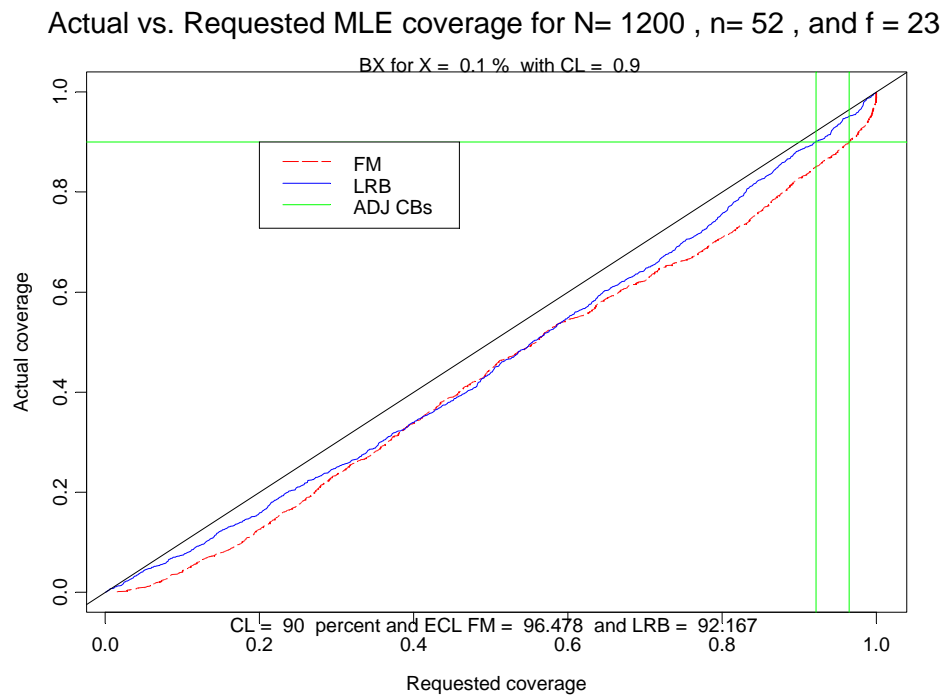


Figure 11. Example graph generated in S-Plus using a modified code that displays the required values and automatically draw the lines.

IV. RESULTS

A. INTRODUCTION

A summary of the relevant features that support the initial propositions of this thesis is included in this chapter, along with a detailed explanation of the simulation output using one of the figures as an example. In addition, a brief discussion of the S-Plus code written for this thesis is presented for users who may need to modify the simulation to include different test sample sizes, different distributions, or different amounts of censoring that were not performed as part of this study. An example is included at the end of this chapter demonstrating the usefulness of the tables and showing the impact of the bias on the acquisition process.

Several of the graphs and tables included in this section show that as the number of censored observations increases (fewer failures), the bias increases. That increase is quantified in Tables 1 and 2 for each of the two methods using a sample size of 10, an 80% confidence level, and varying the number of failures. Based on the results of the simulation study as shown in Tables 1 and 2, the graphical output of the simulation study (Appendix B) was run using the Fisher-Matrix method only, as it consistently produced the greatest amount of bias and is usually the default method used by the listed software packages.

Appendix B is divided into six sections; one section for each reliability value. Each section is sub-divided by the population size (10 and 20). A graph showing the output of the simulation is displayed for each number of failures that occurred during the test period. For example, in Appendix B, Sections A and B contain the results of the simulation study for a contractual 99% reliability. Section A lists the results for a sample size of ten (10) and shows four graphs, one for each of the simulated failures (3, 5, 7, 9), while Section B displays similar graphs, but shows the results of using a sample size of twenty (20).

B. SUMMARY OF GRAPHS

Several observations are evident and worthwhile mentioning in this section. First, by comparing the simulation output of either coverage method in any of the reliability series shown in the body of this thesis, one can see the bias in the confidence interval increases (the simulation value gets farther away from the diagonal line) as the amount of censoring increases (i.e. number of failures in the test sample size decrease from nine to seven to five, etc.). This effect is

graphically displayed in Figure 11 below, where all four of the different failures are plotted simultaneously on one graph for the Fisher-Matrix (FM) method, using a 98% reliability and a sample size of 10. The most error in the confidence bounds occurred when the data is heavily censored (i.e., only three failures and seven still working).

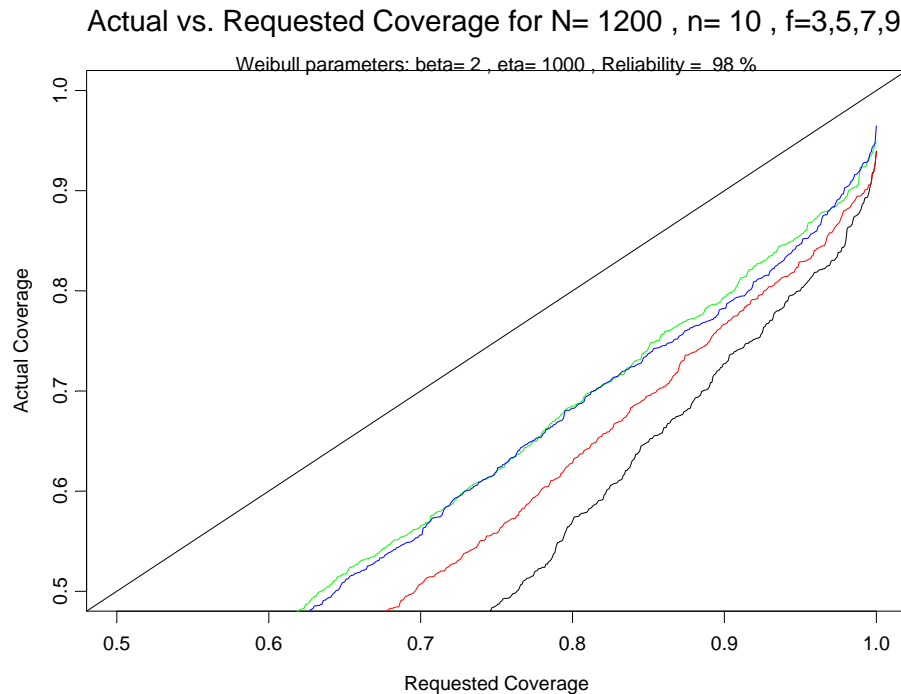


Figure 12. A graph showing an increasing error in the confidence bounds as the numbers of failures decrease for the Fisher-Matrix Method for 98% reliability and a test population of 10.

Note: The data is the result of 1200 Monte Carlo simulations using S-Plus with SPLIDA. The colors are in order of closest to the diagonal line to farthest away. (green = 9 failures, blue = 7 failures, red = 5 failures, black = 3 failures)

The second noticeable finding in this simulation study is that the Fisher-Matrix method of determining confidence bounds has more bias than the Likelihood Ratio Bounds method (LRB). The small sample size coupled with censored data has more of an effect on the FM method than the LRB method and supports the theorists who have asserted this point for years. Tables 1 and 2 quantify the bias for each of the two methods using a sample size of 10, an 80% confidence level, and varying the number of failures.

Based on the results of the simulation study as shown in Tables 1 and 2, the graphical output (Appendix B) of the simulation study only was run using the FM method as it consistently produced the greatest amount of bias. The following tables are based on 10000 iterations from using the Monte Carlo code written for this thesis.

| Number of Failures | Reliability | | | | | |
|--------------------------|-------------|-------|-------|-------|-------|-------|
| | 99% | 98% | 97% | 96% | 95% | 90% |
| 3 | 20% | 18.8% | 18.8% | 16.3% | 16.3% | 5% |
| 5 | 17.5% | 16.3% | 13.8% | 13.8% | 13.8% | 11.3% |
| 7 | 15% | 13.8% | 12.5% | 13.8% | 12.5% | 10% |
| 9 | 13.8% | 12.5% | 11.3% | 13.8% | 13.8% | 11.3% |

Table 1. Percent error in the confidence bounds using the Fisher-Matrix Method (FM), a sample size of 10, and an 80% (actual) confidence level. (Where the Error = $\text{abs}[(\text{Required}-\text{Actual})/\text{Actual}] \times 100\%$).

| Number of Failures | Reliability | | | | | |
|--------------------------|-------------|-------|-------|-------|-------|-------|
| | 99% | 98% | 97% | 96% | 95% | 90% |
| 3 | 13.8% | 12.5% | 13.8% | 8.8% | 8.8% | 0% |
| 5 | 11.3% | 10% | 11.3% | 8.8% | 11.3% | 6.3% |
| 7 | 8.8% | 11.3% | 11.3% | 7.5% | 6.3% | 6.3% |
| 9 | 6.3% | 8.8% | 19% | 11.3% | 10% | 11.3% |

Table 2. Percent error in the confidence bounds using the Likelihood Ratio Bounds (LRB) method, a sample size of 10, and an 80% (actual) confidence level. (Where the Error = $\text{abs}[(\text{Required}-\text{Actual})/\text{Actual}] \times 100\%$).

C. ERROR CORRECTION USING THE CATALOG

The following tables summarize the required value for the FM method that one must specify in order to obtain an 80% confidence bound for each of the simulations in this thesis. These values can be read directly from the graphs in Appendix B or can be produced by the simulation code running the subroutine `get.hartley.lcb'` and are representative of the information available in the Appendix. The 80% confidence bound was chosen as a demonstration of the information available from the catalog.

Users of this catalog must specify the “Requested” value that is read from the appropriate graph that meets their particular needs and meets the small sample criteria with censored data. So for a 97% reliability and an 80% confidence for a sample size of 10 with 5 failures, one would have to specify an 89% confidence level in their software in order to meet a true 80% confidence as required by the specification.

| Number of Failures | Reliability | | | | | |
|--------------------|-------------|-------|-------|-------|-------|-------|
| | 99% | 98% | 97% | 96% | 95% | 90% |
| 3 | 99.99 | 99.98 | 99.98 | 99.97 | 99.97 | 99.7 |
| 5 | 99.04 | 98.69 | 98.68 | 98.17 | 98.16 | 97.54 |
| 7 | 96.84 | 96.40 | 95.94 | 95.71 | 96.06 | 94.60 |
| 9 | 94.20 | 93.65 | 94.14 | 93.76 | 93.81 | 93.09 |

Table 3. Summary of the required confidence using the Fisher-Matrix method (FM) and an actual 80% confidence using a Monte Carlo simulation with a sample size of 10 with 3, 5, 7, & 9 failures for 99%, 98%, 97%, 96%, 95%, and 90% reliability.

| Number of Failures | Reliability | | | | | |
|--------------------|-------------|-------|-------|-------|-------|-------|
| | 99% | 98% | 97% | 96% | 95% | 90% |
| 3 | 99.99 | 99.98 | 99.97 | 99.91 | 99.80 | 95.83 |
| 5 | 99.02 | 98.63 | 98.68 | 98.41 | 97.91 | 95.28 |
| 7 | 97.27 | 96.48 | 96.48 | 96.40 | 96.18 | 94.27 |
| 9 | 95.40 | 94.85 | 94.67 | 93.86 | 93.85 | 92.25 |

Table 4. Summary of the required confidence using the Fisher-Matrix method (FM) and an actual 80% confidence using a Monte Carlo simulation with a sample size of 20 with 3, 5, 7, & 9 failures for 99%, 98%, 97%, 96%, 95%, and 90% reliability.

Although Table 4 above shows values in the 90% reliability column for both three and five number of failures, the values are questionable to a degree. The graphs of the simulation for these two failures (Appendix B, Section L) show an inflection point in the output data where the graph rises above the diagonal line (lower half). This appears to be a result of the heavily censored data (only 3 of 20 and 4 of 20 failures) coupled with a low reliability (90%). The numbers appear to follow the trends in Table 4 and the upper half of the graphs in Appendix B, Section L are of similar magnitude as other graphs in that class, therefore I left the entries in

Table 4 and suggest they are reasonable for this thesis. Clearly, however, small numbers of failures result in inaccurate coverage probabilities for low reliabilities. This behavior has been confirmed by repeated trials and warrants further study.

D. CODE MODIFICATION FOR SPECIFIC APPLICATIONS

The S-Plus code that was used in this thesis can be modified to change many of the parameters used to describe the life data analysis. In particular, the sample size and number of failures can be specifically tailored to accommodate many users.

The first line of the code is the ‘function’ line and contains all of the relevant parameters including number of iterations (N), sample size (n), and number of failures (f), and confidence level (g). The function line appears below as used in this thesis.

```
get.hartley.lcb <- function(N = 10000, g = 0.95, n = 13, f = 10, alpha = 0.1, plot.it = F)
```

E. TEST CASE

To demonstrate an application of the usefulness of the tables in this thesis and as another example of modifying the S-Plus simulation code, this section will replicate the results of a worked example in Lawless’ text on lifetime data [Ref; 10, p. 155]. In his text, he worked an example using failure times from a small sample of aircraft components with censored data. His solution was based on the Weibull distribution and used tables that appeared in a technical report [Ref: 11] from Wright-Patterson AFB, OH to obtain a value for 90% reliability and a 95% lower, one-sided confidence bound. The exact value of B10C95 obtained using the tables in Reference 11, was (.105).

The data shown in the table below are the 10 failure times of aircraft components and the three censored times all stopped after the tenth failure. The values of $N = 13$ and $f = 10$ were used as inputs to the S-Plus simulation code used in this thesis. The function line of the code is shown below:

function(N = 10000, n = 13, f = 10, alpha = 0.10, plot.it = T)

| Failure (F)/Censored(S) | Time (hours) |
|-------------------------|--------------|
| F | 0.22 |
| F | 0.5 |
| F | 0.88 |
| F | 1 |
| F | 1.24 |
| F | 1.33 |
| F | 1.54 |
| F | 1.76 |
| F | 2.5 |
| F | 3 |
| S | 3 |
| S | 3 |
| S | 3 |

Table 5. Aircraft component failure times in hours. [Ref:10, p. 155]

The output of the simulation code for the FM method is shown in Figure 12. Using the same graphical technique that was described in Figure 9, the ‘required confidence’ was located by drawing a horizontal line from the y-axis at the confidence value your specification calls for (in this case 95%), and intersecting with the simulation output. Then drawing a vertical line down to the x-axis and estimating the new confidence value that you need to request from the contractor or testing organization (approximately 99.5%).

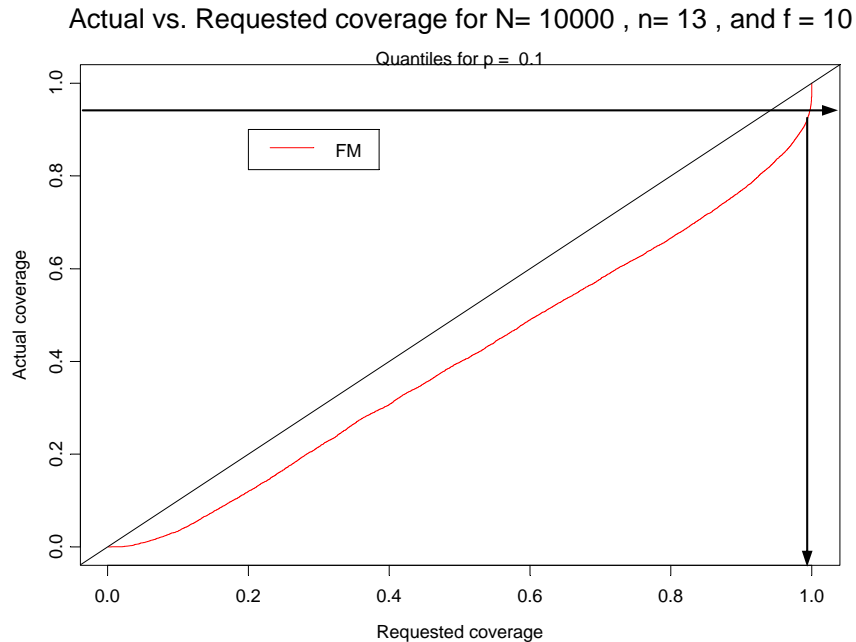


Figure 13. Output of the S-Plus simulation code using $N=13$ and $f=10$ derived from Example 4.1.1. [Ref: 10]

An output of the S-Plus simulation code subroutine 'get.hartley.lcb' is shown below. The first line matches the exact value of B10C95 as calculated by Lawless [Ref: 10, p. 155] and indicates the simulation code is working correctly. The following lines are the final lines of the S-Plus output and specifies the 'required' confidence directly without using the graphs.

The true 0.1 point is 0.1053

We have a sample of size 13 and 10 failures.

To get a 95 percent actual confidence level, you must request a 99.85 percent FM confidence level.

The results of this simulation output are close to the estimated value of the required confidence using the visual aid of drawing the line across and down from Figure 13 as shown above (approximately 99.5%). The simulation output may provide a more accurate confidence value for those who have S-Plus and SPLIDA as analysis tools but the graph technique using the graphs in Appendix B is also quite good.

As a final verification of the simulation code and another indication that the ‘required’ confidence from the simulation is credible, I analyzed the Lawless data in another life-data analysis program called Weibull++6 [Ref. 12]. In particular, I was interested in obtaining the B10C95 (shorthand for the 95% lower confidence bound for the tenth percentile) value using both the original confidence value (95%) , and then recalculating using the value obtained from the S-Plus output (either from the graphs or the direct readout from the subroutine ‘get.hartley.lcb’).

Using the data from Table 5, the Weibull++6 program provided the following graph shown in Figure 14, as well as Tables 6 and 7.

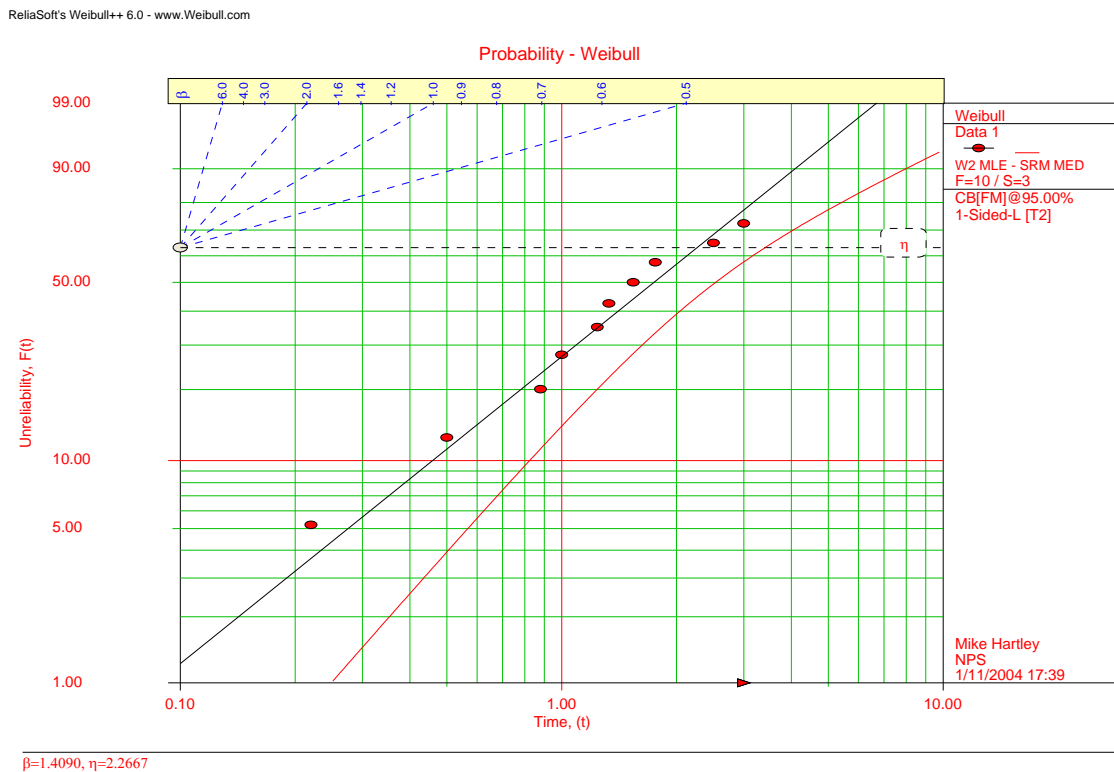


Figure 14. The output of the Weibull++6 software package used to determine the B10C95 value for the S-Plus simulation code.

| | |
|---------------------------|---------------|
| User Input: | |
| BX% information at: = | 10 |
| Confidence Bounds Used: | 1-Sided |
| Confidence Bounds Method: | Fisher Matrix |
| Confidence Level: = | 0.95 |
| Weibull++ Output : | |
| Time: = | 0.459 |
| Lower Limit: = | 0.205 |
| Confidence: = | 1S @ 0.95 |

Table 6. Output from the Weibull++6 Life Data Analysis Software package used for comparison with Example 4.1.1. [Ref: 10, p. 155]

Table 6 shows that the lower limit using the Fisher-Matrix method was (.205) compared to the exact value of (.105) obtained by Lawless [Ref: 10, p. 155]. The B10C95 value using the original confidence value of 95% is almost twice the exact value calculated by Lawless!

The ‘required’ confidence value of 99.85% from the subroutine ‘get.hartley.lcb’ was then used in the Weibull++6 software package to determine a new lower limit. Table 7 displays the result of the calculations and shows the lower limit of (.1073) using the estimated value from the simulation code for the confidence value. The B10C95 now compares favorably to the exact value of (.105) as determined by Lawless.

| | |
|---------------------------|---------------|
| User Input: | |
| BX% information at: = | 10 |
| Confidence Bounds Used: | 1-Sided |
| Confidence Bounds Method: | Fisher Matrix |
| Confidence Level: = | 0.9985 |
| Weibull++ Output : | |
| Time: = | 0.459 |
| Lower Limit: = | 0.1073 |
| Confidence: = | 1S @ 0.9985 |

Table 7. Output from the Weibull++6 Life Data Analysis Software package using the confidence value from the S-Plus simulation code for Example 4.1.1. [Ref: 10, p. 155]

V. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The objective of this thesis was to graphically show that analyzing small sample sizes, commonly found in most Developmental Test (DT) scenarios, with software designed for large samples, impacts many areas in the acquisition analysis and the *small* error that has been acknowledged for years has a *large* impact in today's economy. The results of the simulation study in this thesis quantify the bias in one-sided reliability confidence bounds when using small samples and censored data.

Tables 1 and 2 in the previous section are tabular displays of the error in the confidence bounds and show that for this simulation study the error can be as much as 20%. The results section conclusively shows that the Likelihood Ratio Bounds method (13.8% error) has less bias than the more commonly used Fisher-Matrix method (20% error) especially for high reliability when the number of failures is small.

The test case shown in the previous section also illustrates the impact of small sample size analysis using actual aircraft component data. The comparison of the B10C95 using the original confidence value in the Weibull++6 software package, to the exact value calculated by Lawless shows almost a 100% increase (.105 versus .205). This increase can directly translate into overestimating inspection times for mission critical components and purchasing and storing of unnecessary spares, increasing the logistic tail for combat deployments.

The bias can be reduced when using standard software packages to analyze small sample with censored data by either the graphical technique using the graphs in Appendix B, or running the simulation code using S-Plus with the embedded SPLIDA life data routine. The test example showed that using either technique can reduce the error from 95.23% (using the 95% confidence) to less than 1.9% (using the 'required' 99.85% confidence obtained from Figure 13 and the output of 'get.hartley.lcb').

Although increasing the number of items under test can increase the confidence, this thesis shows that the major influence is the small number of failures that occur during the test time. Many of the governments' contracts specify quantities such as reliability and confidence

but we can't mandate the number of test items because of the constraint to minimize costs and the ever-present desire to shorten the test time. An unknown author writes what is known as The Paradox of Reliability Analysis,

The more reliable a product is, the harder it is to **get** the failure data needed to "prove" it is reliable!

However, we can now argue for better analysis by understanding the bias associated with small samples with censored data and modifying our requirements to account for the error.

B. RECOMMENDATIONS

Two improvements to the simulation code appeared during the writing of this thesis. The first and most important improvement would be to modify the simulation code for other life data analysis software packages. Although the use of S-Plus is widespread in industry, many other analysis packages exist and are used worldwide. The second improvement would be to write a simulation code for other high-speed computer software programs that are not strictly life data analysis tools. For example, Dean M. Ford from the Delphi Automotive Systems has developed a simulation code for the Weibull distribution that runs on Excel and is much faster than the S-Plus code used in this thesis.

Finally, in addition to improving the speed of the code used in this thesis, this code could easily be extended to include other distributions such as exponential, lognormal and gamma. An efficient simulation tool or catalog listing all of the location-scale distributions showing the required confidence for small samples would be an invaluable tool for anyone dealing with life data analysis, reliability, and warranty contracts.

APPENDIX A

This section contains the S-Plus code that was written for this thesis. It performs a Monte Carlo simulation for the Weibull function using the number of test items and the number of failed units. This code can be copied and used in the function command in S-Plus. After naming the function, run it by calling the function and specifying the number of iterations in parentheses similar to the following line: `> get.hartley.lcb(N = 10000)`. If you require a specific confidence level and would like the S-Plus program to calculate determine that value for you, modify the `'get.hartley.lcb` code in the function line below by specifying the reliability (alpha), and save `'get.hartley.lcb`. Now run the function `'get.hartley.lcb` and set the variable `'g` to the confidence you want. The program will calculate the required confidence level and run the function resulting in a plot of the actual versus required confidence for all confidence values. (Note: The SPLIDA subroutine must be loaded into the S-Plus code in order for this code to work.)

```
#####
# This program is the final working version of Mike Hartley's
# routine to get equivalent confidence levels for BXLBYYY
#
# to use, load the first function. A typical function call is the
# last line of the code below.
#
# 2/29/04; DHO
#
# "Team players use S Plus, but it takes the JV longer...."
#
#####

get.hartley.lcb<-function(N=10000, g=.95, n=13, f=10, alpha = 0.1, plot.it=F)
{
  #bad parameter traps
  if(alpha > 0.5)
    stop("Try another alpha. \n \n")

  if(g>=1 || g<=0) stop ("Try another g. \n \n")
}
```

```

true.pt<-exp(qsev(alpha)) # the true percentile
cat("The true", alpha, "point is", true.pt, "\n")

test.out<-SingleDistSim(number.sim=N,"Weibull",sample.size=n,
  censor.type = "Type 2", fail.number = f, kprint = 0) #gets a sample from the weibull of
size N

mu<-test.out$theta.hat[,1]
sigma<-test.out$theta.hat[,2]
var.mu<-test.out$vcvobs[,1]
var.sigma<-test.out$vcvobs[,3]
covar<-test.out$vcvobs[,2]

tp<-exp(mu + qsev(alpha)*sigma) # tp is the vector of point estimates from the
samples

se.tp<- tp *sqrt(var.mu + 2* qsev(alpha) *covar +(qsev(alpha))^2 * var.sigma) #the
vector of SEs for tp

ecl<-(pnorm(tp*log(tp/true.pt)/se.tp)) #the CL necessary to cover the true value for each
sample

ecl<-sort(ecl) #the sorted ECLs

if (plot.it==T){

  plot(ecl,((1:N)-.5)/N, type='n', xlab='Requested coverage', ylim=c(0,1), ylab='Actual
coverage', main = paste("Actual vs. Requested coverage for N=", N, ", n=", n, ", and f=", f))
  lines(ecl,((1:N)-.5)/N, col=2)
  mtext(paste("Quantiles for p = ",alpha))
  abline(0,1,col=1)
  legend(0.2,0.9,c("FM"), lty=c(1,1),col=c(2,2))
}

cat(paste("We have a sample of size", n, " and ", f, " failures. \n"))

cat(paste("To get a", 100*g, " percent Hartley confidence level, you must request a",
  100*quantile(ecl,g), "percent FM confidence level. \n"))

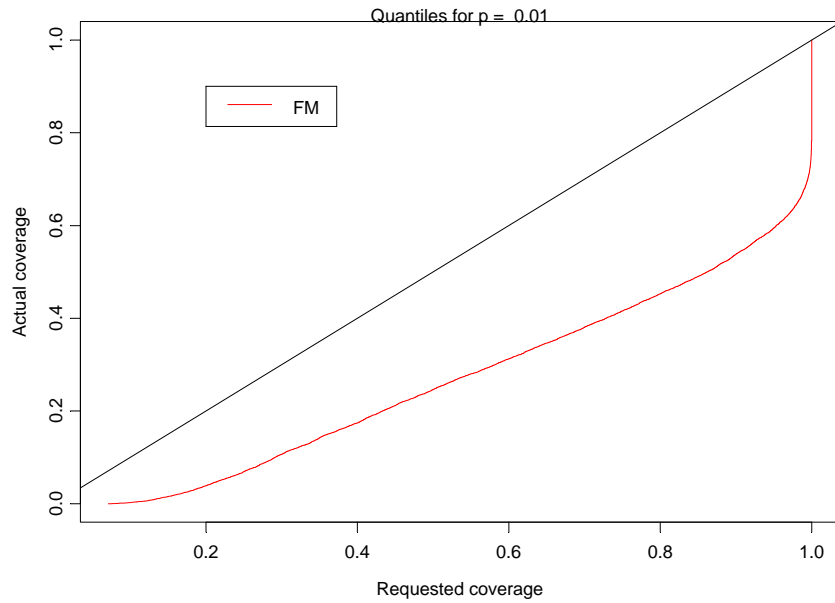
}
get.hartley.lcb(N=10000, n=13, f=10, alpha = .1, g = .95, plot.it = T)

```

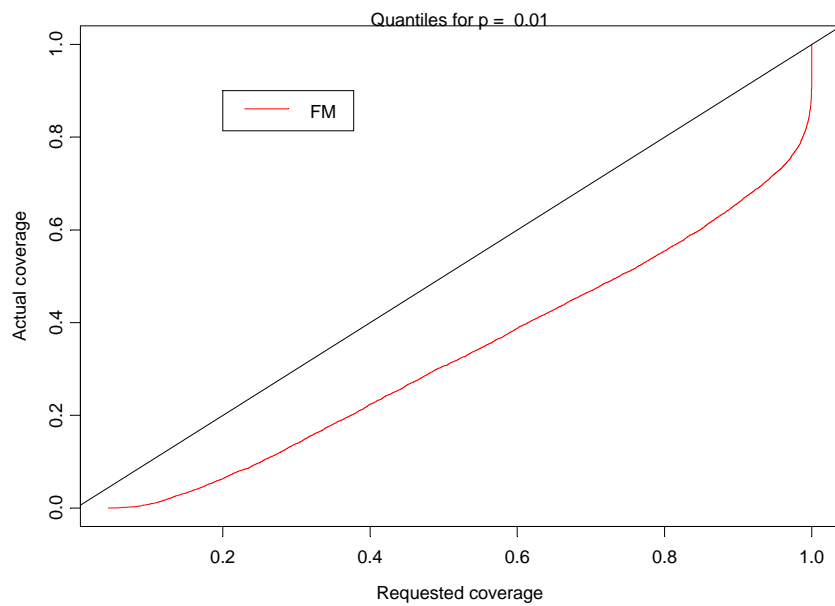
APPENDIX B

A. RELIABILITY = 99%, SAMPLE SIZE = 10

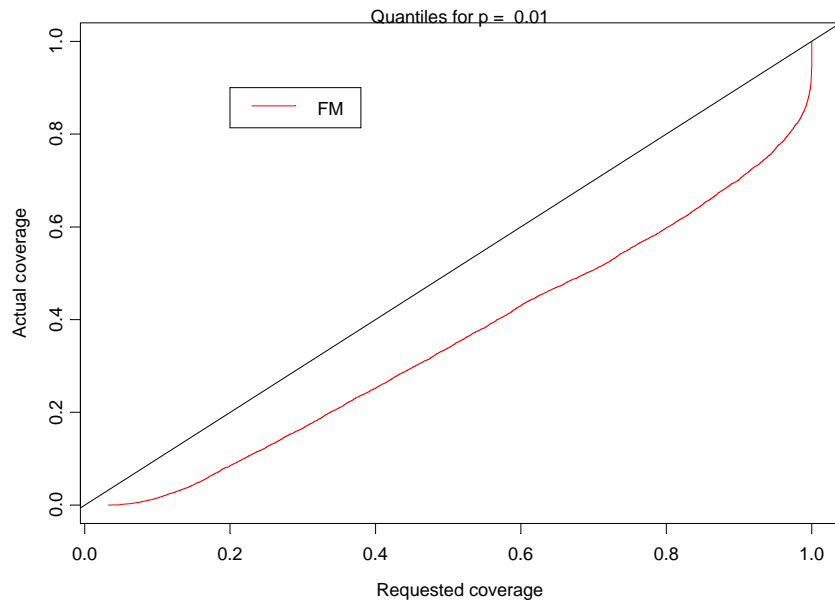
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 3$



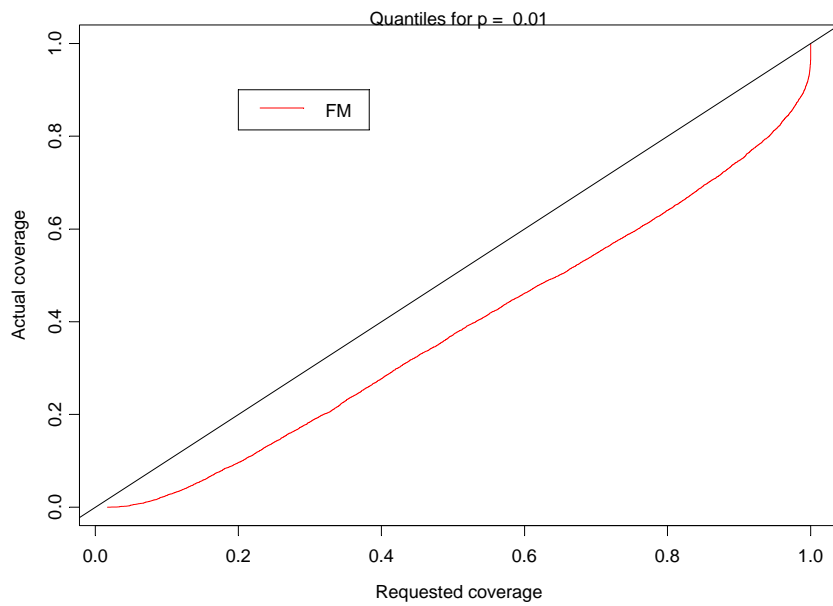
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 5$



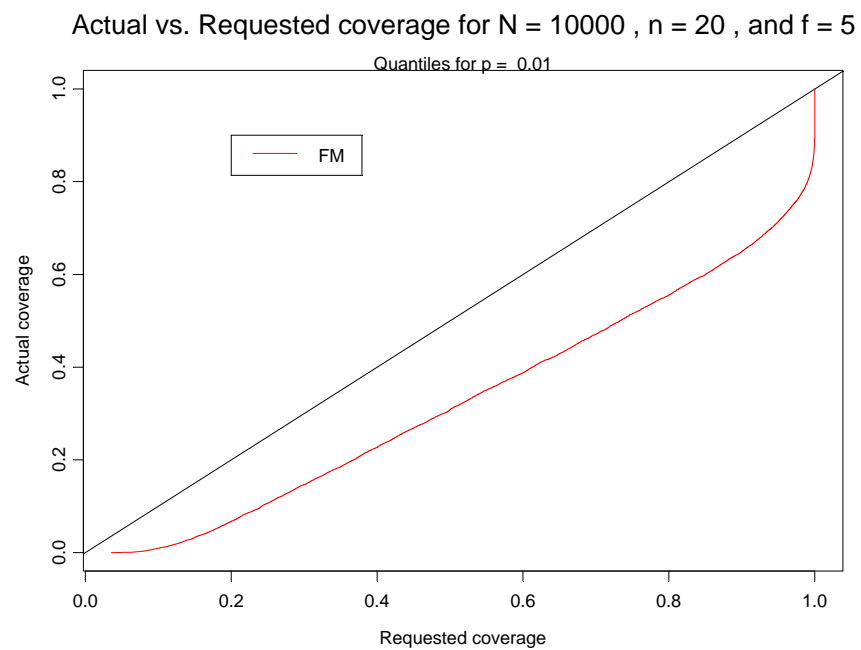
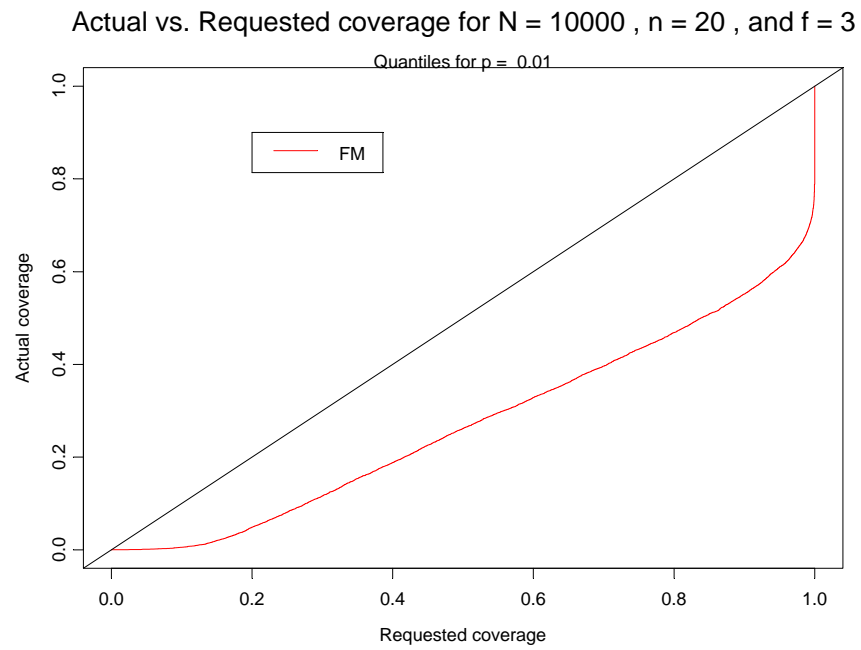
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 7$



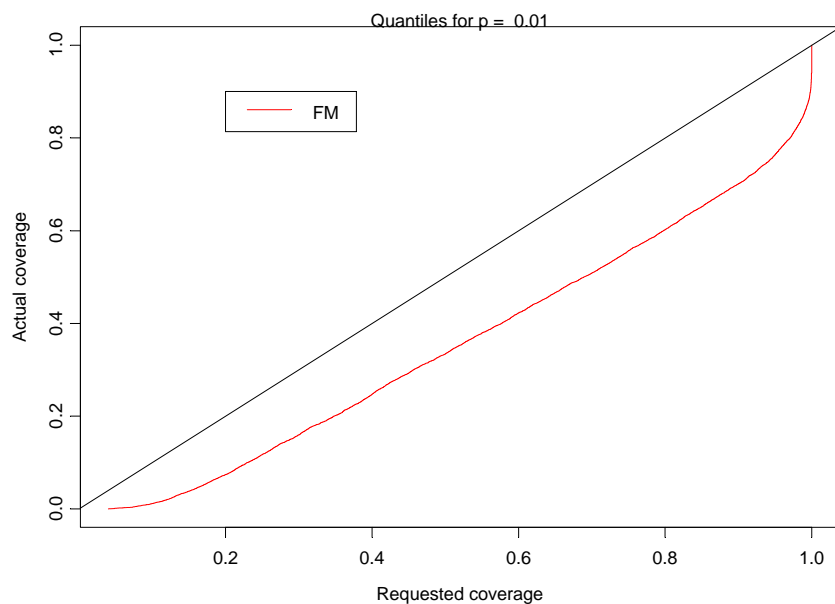
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 9$



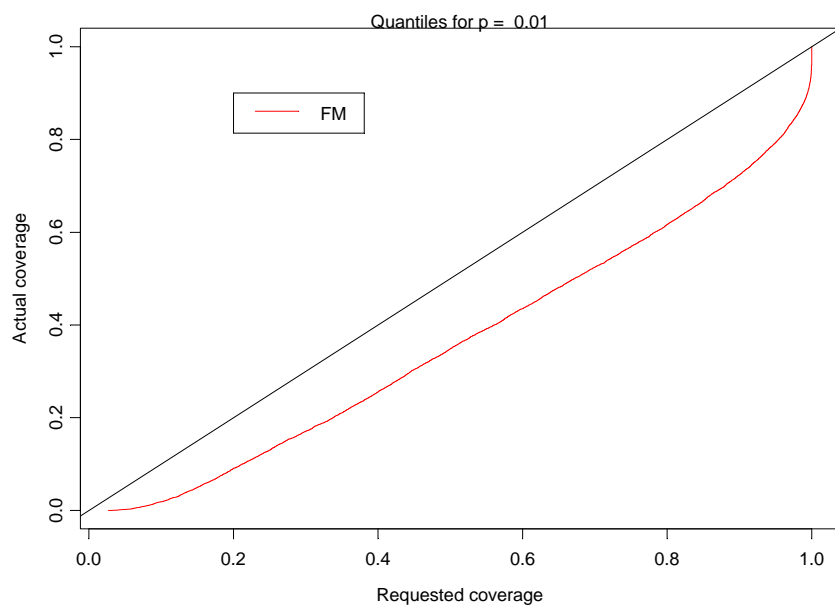
B. RELIABILITY = 99%, SAMPLE SIZE = 20



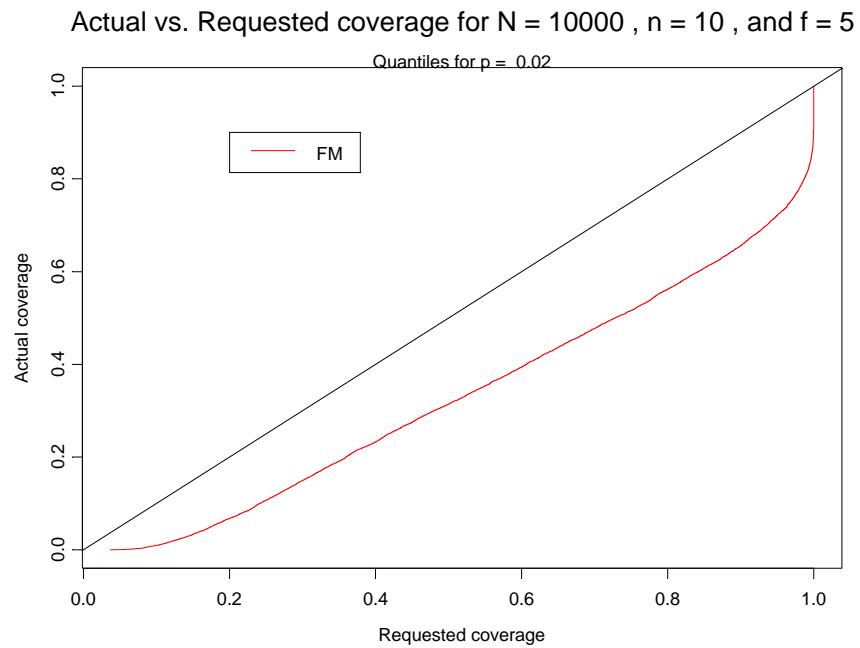
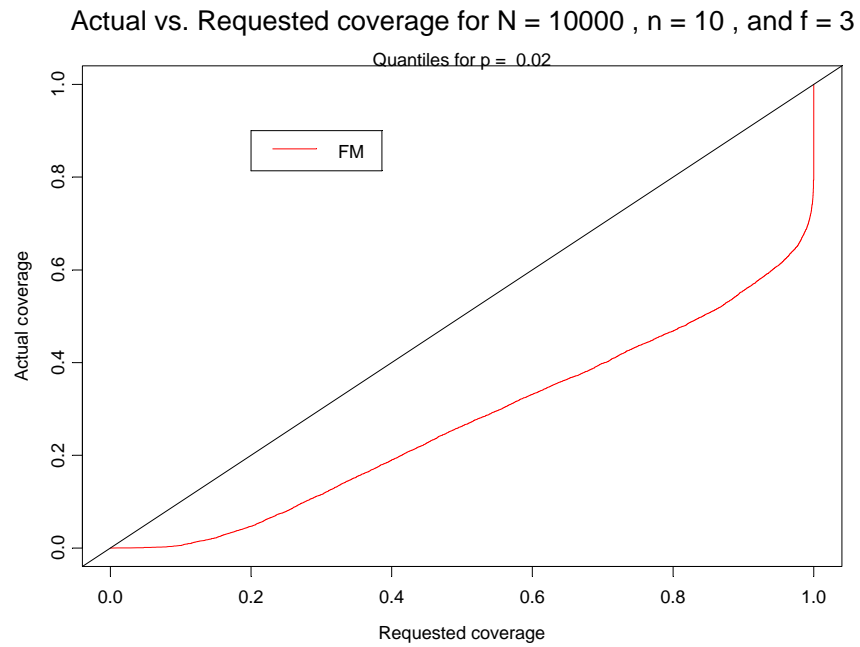
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 7$



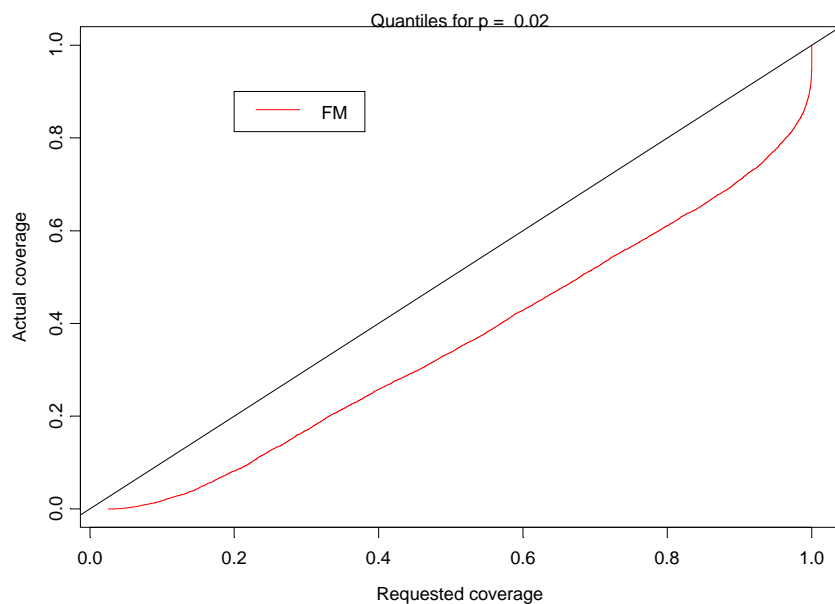
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 9$



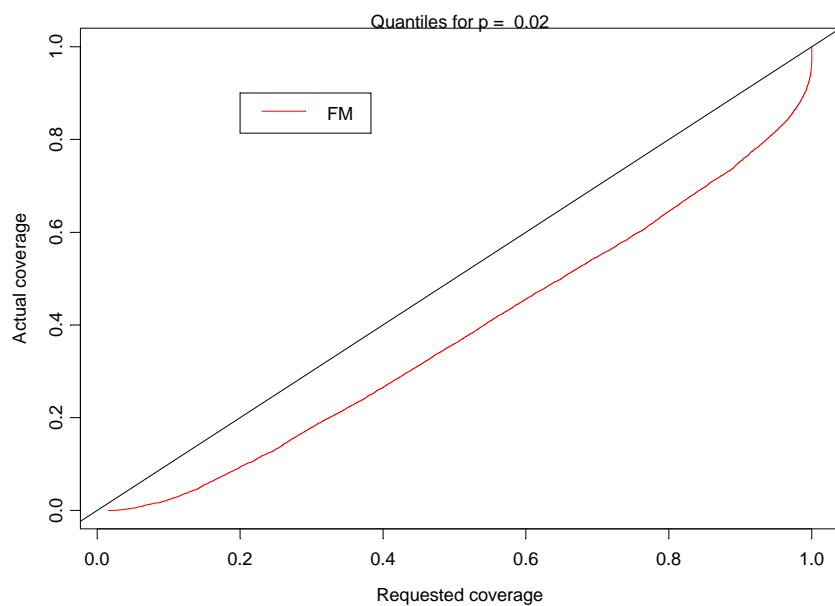
C. RELIABILITY = 98%, SAMPLE SIZE = 10



Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 7$

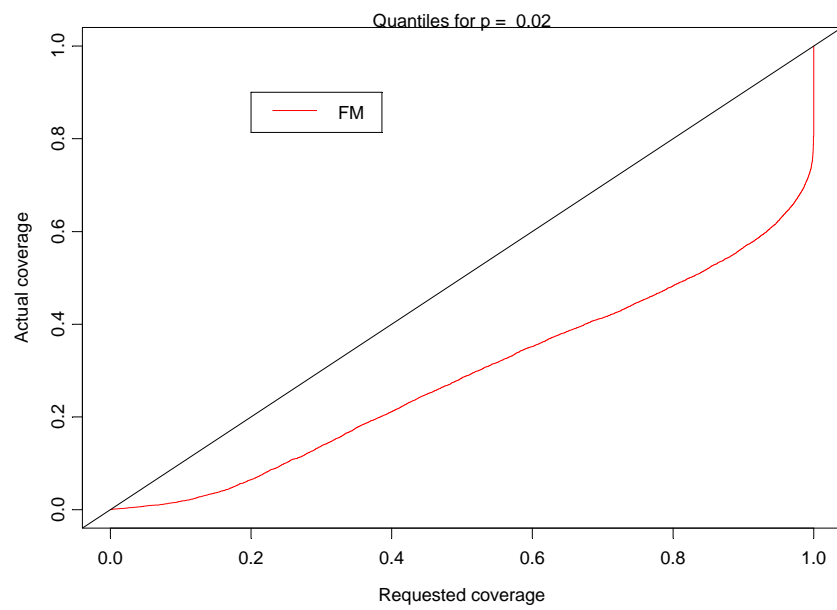


Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 9$

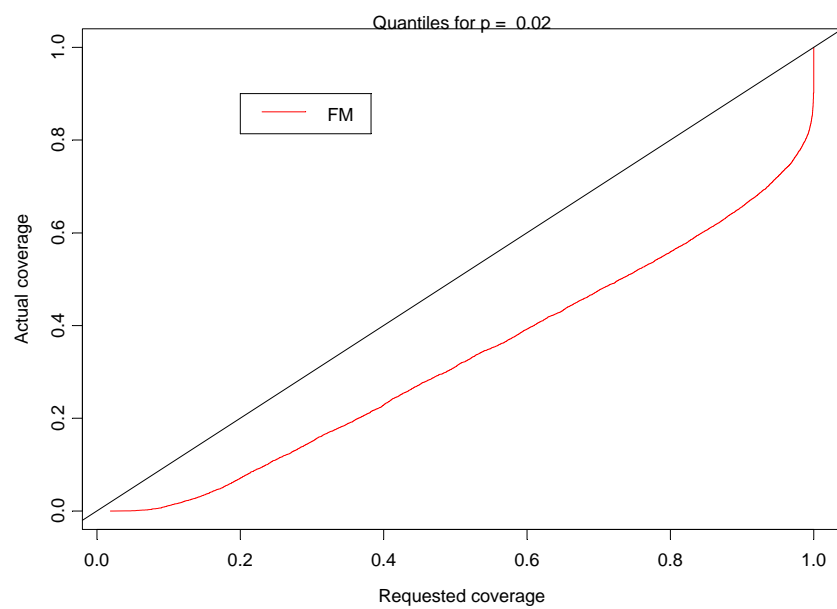


D. RELIABILITY = 98%, SAMPLE SIZE = 20

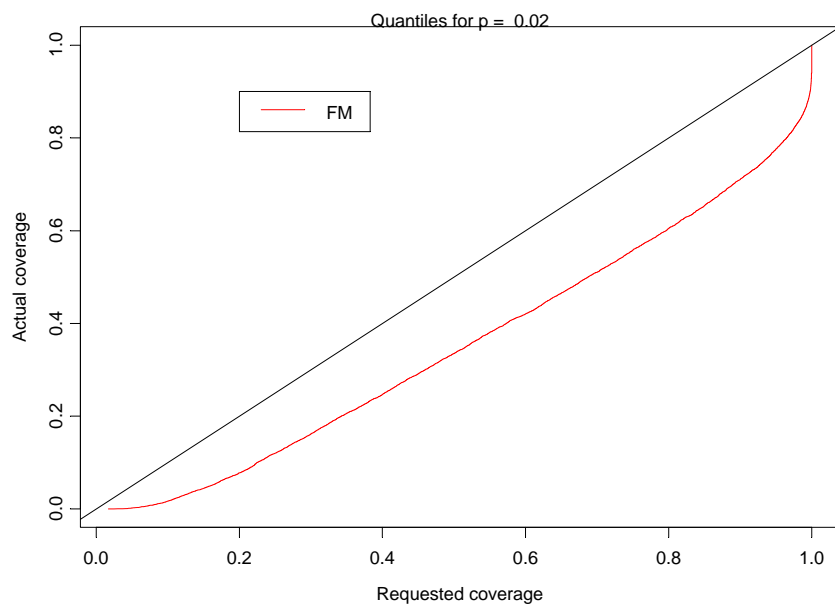
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 3$



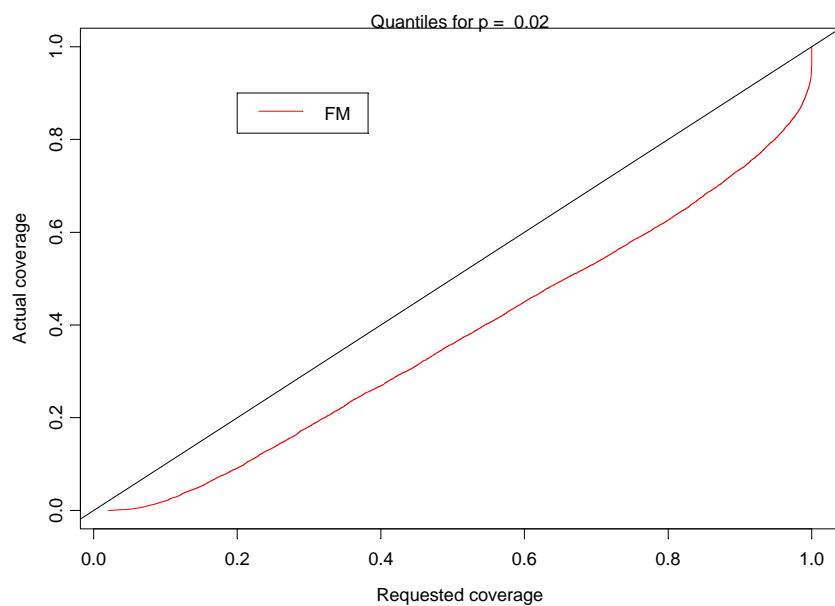
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 5$



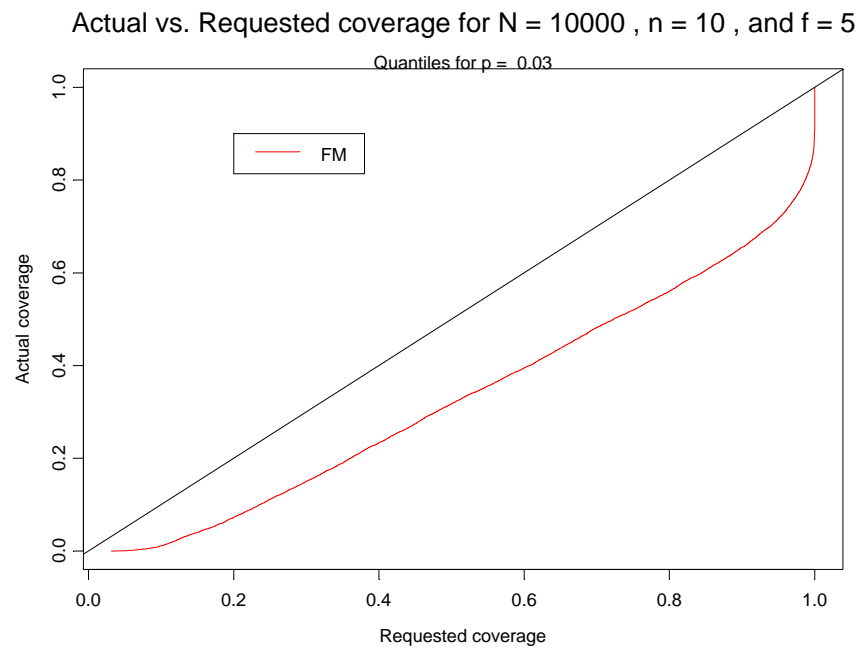
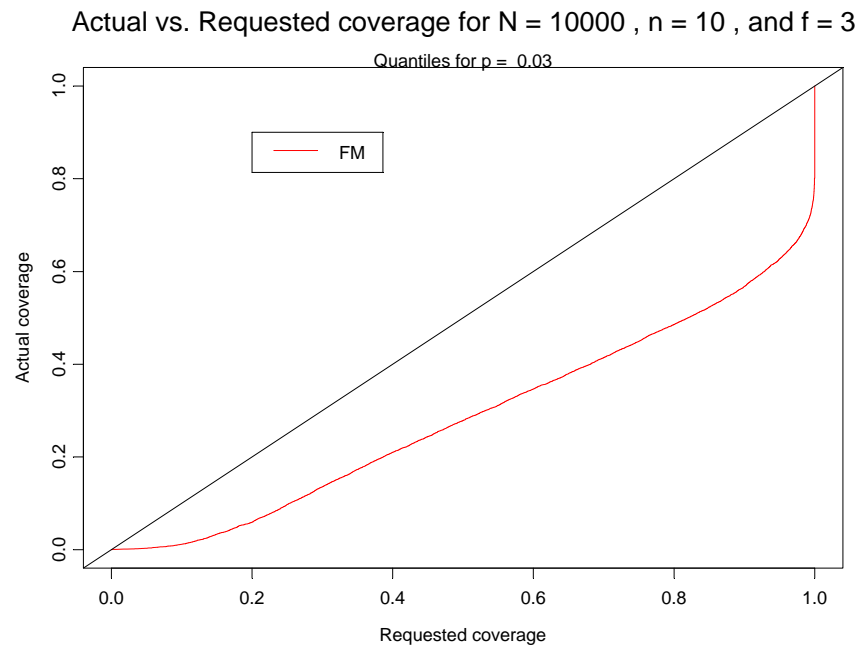
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 7$



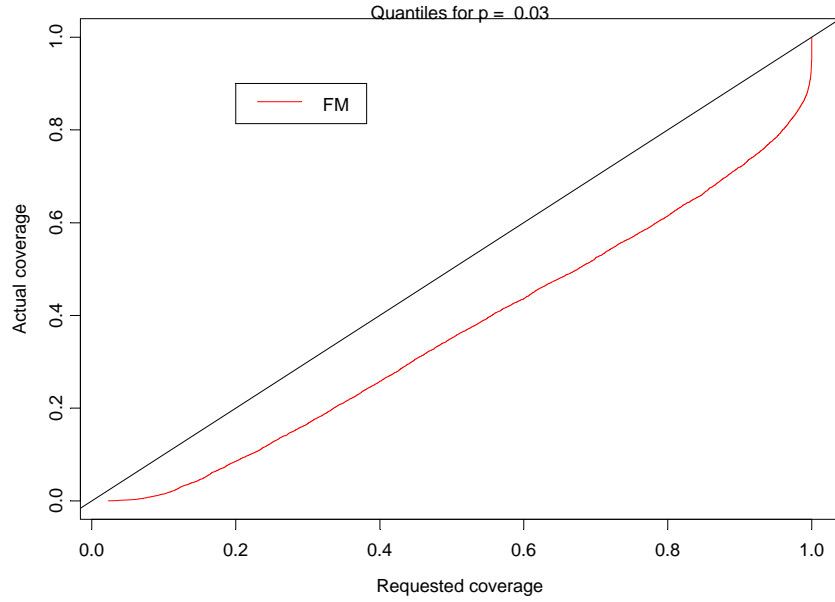
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 9$



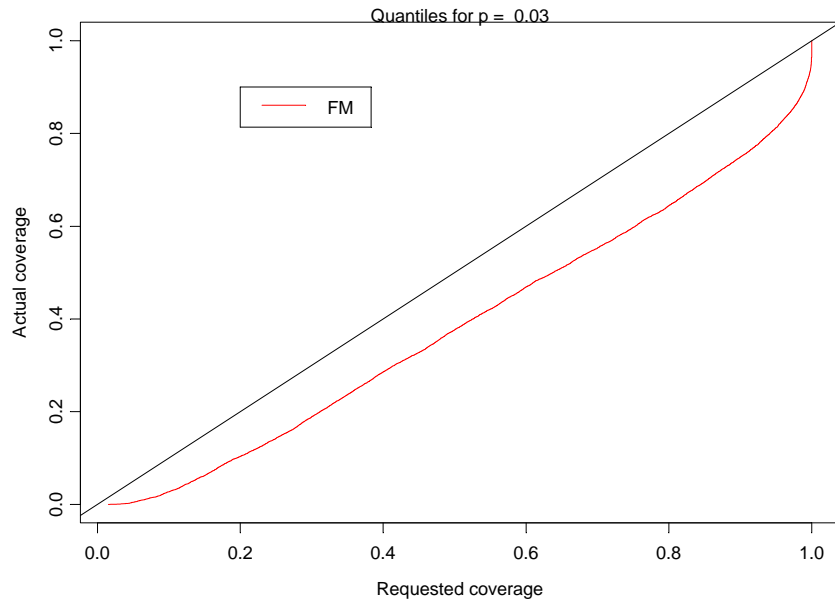
E. RELIABILITY = 97%, SAMPLE SIZE = 10



Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 7$

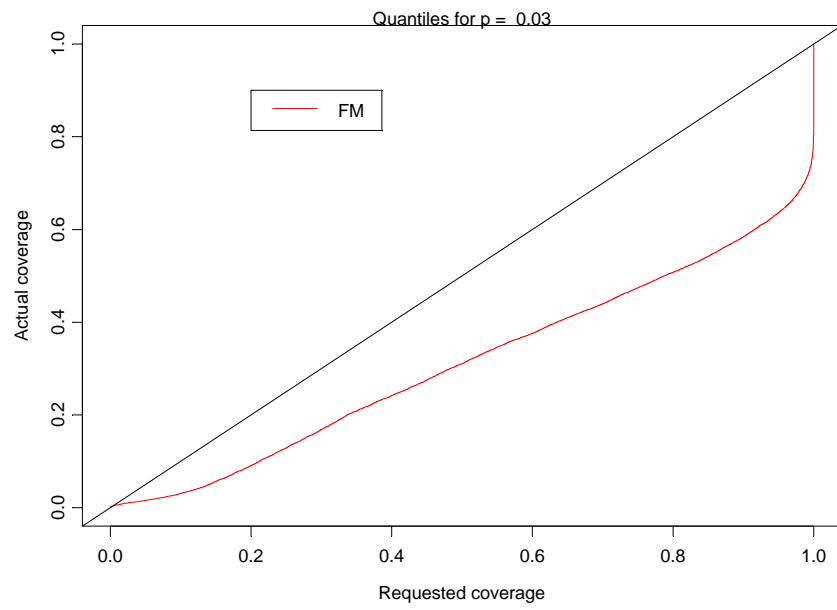


Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 9$

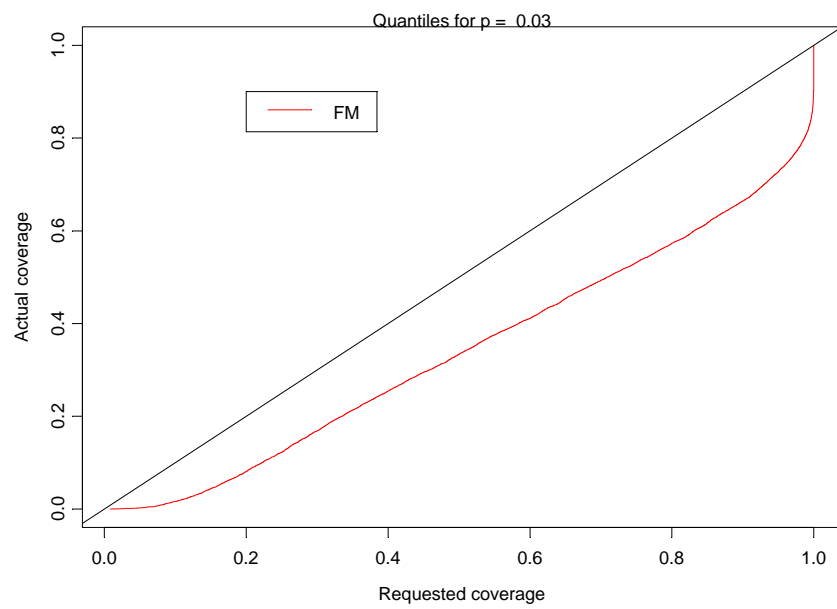


F. RELIABILITY = 97%, SAMPLE SIZE = 20

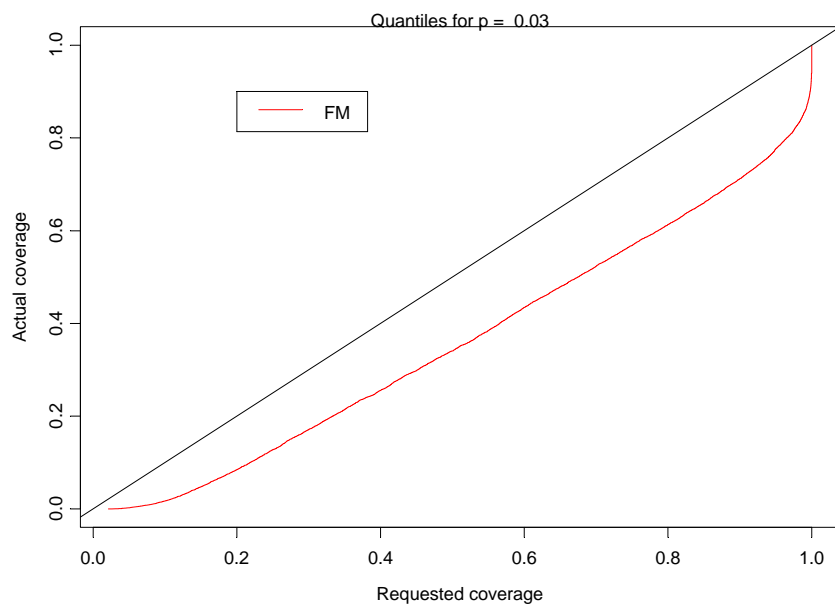
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 3$



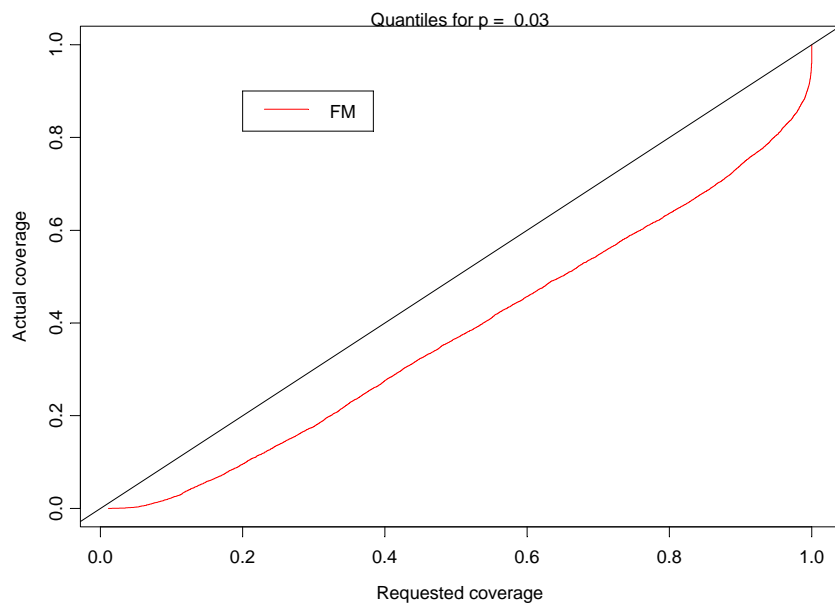
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 5$



Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 7$

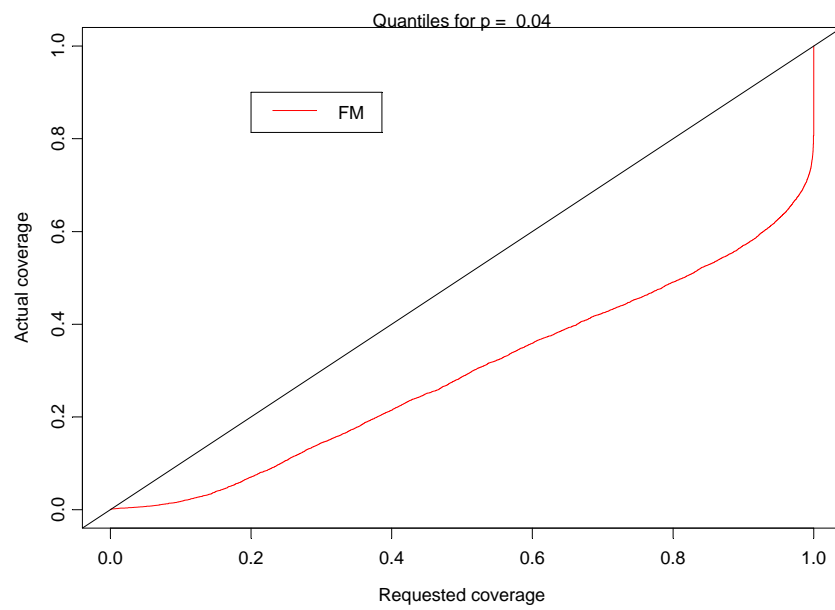


Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 9$

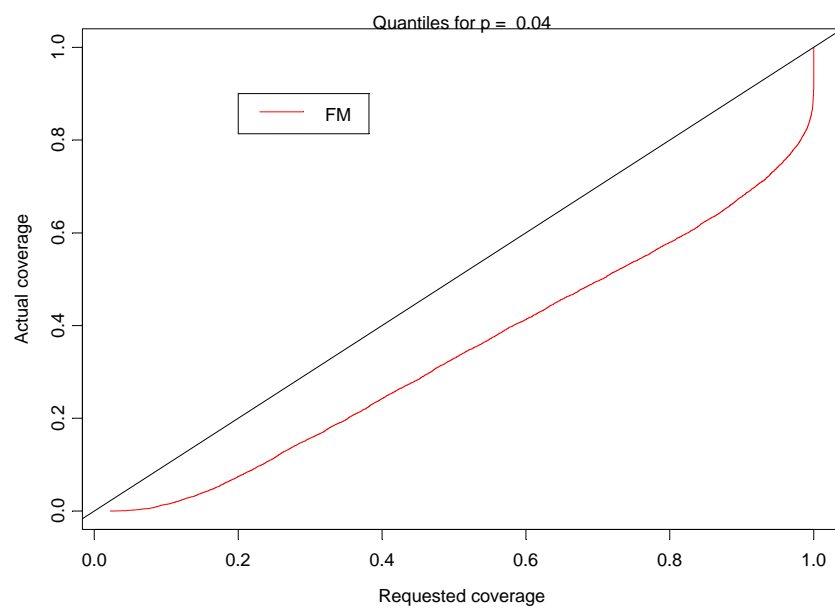


G. RELIABILITY = 96%, SAMPLE SIZE = 10

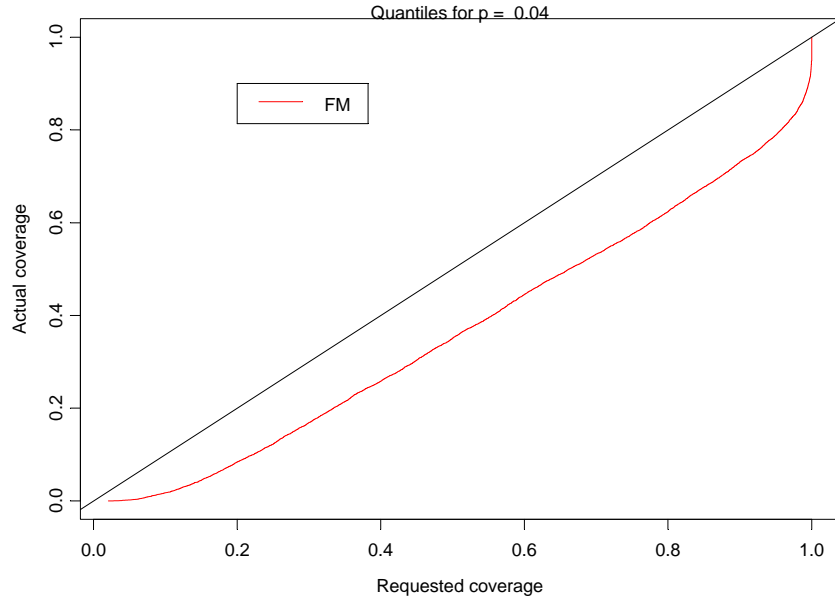
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 3$



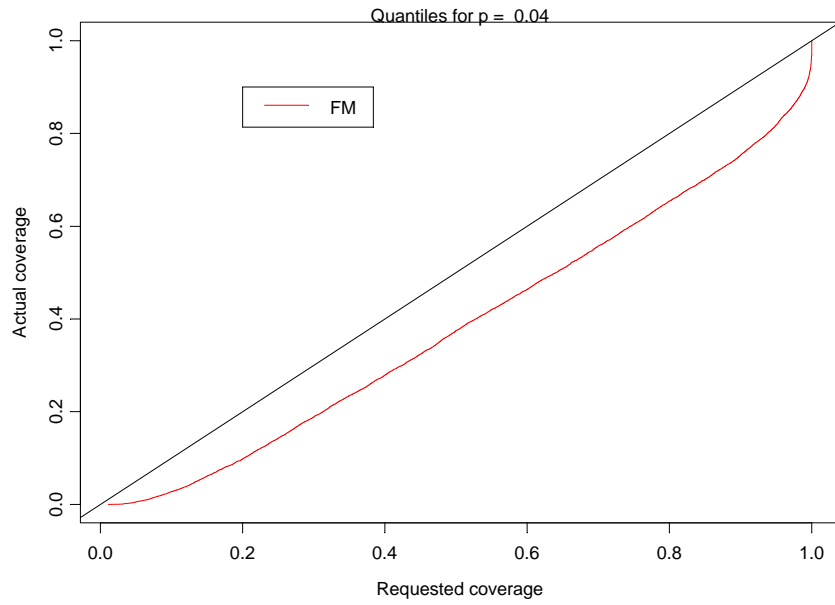
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 5$



Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 7$

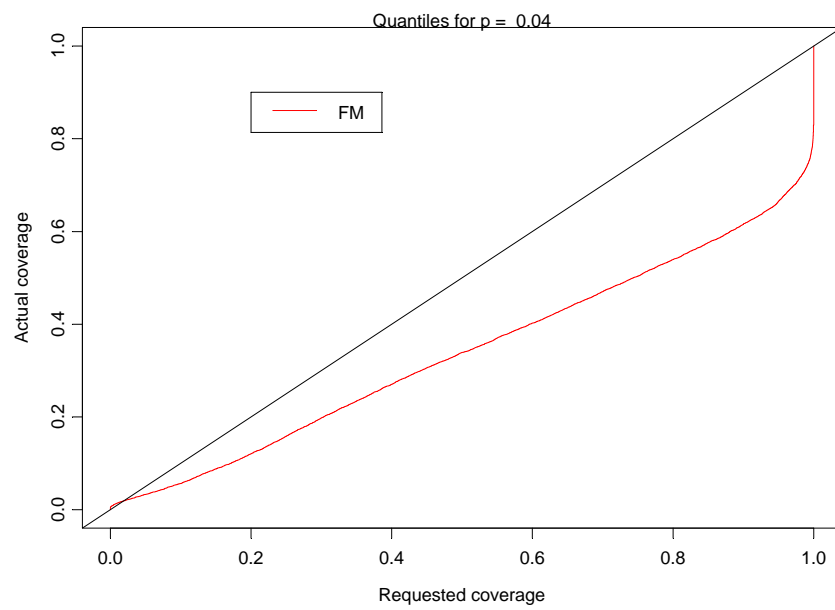


Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 9$

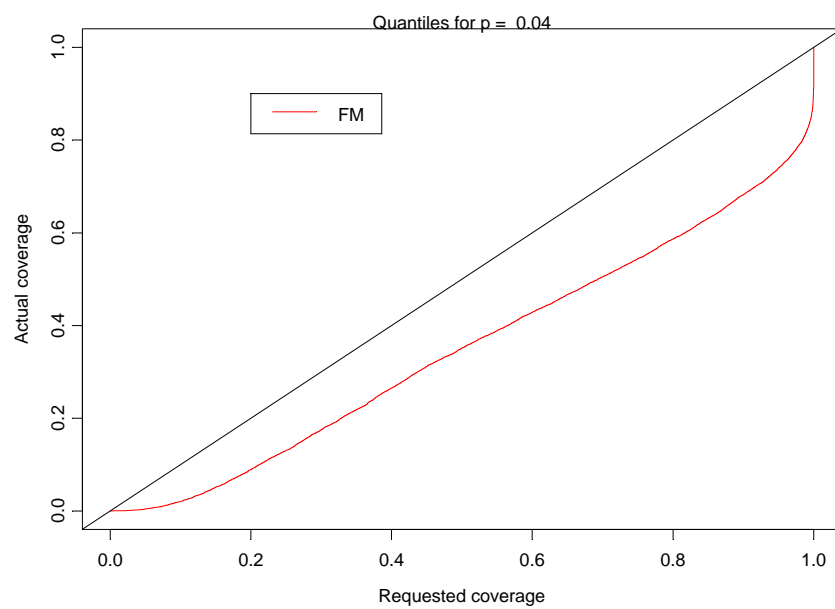


H. RELIABILITY = 96%, SAMPLE SIZE = 20

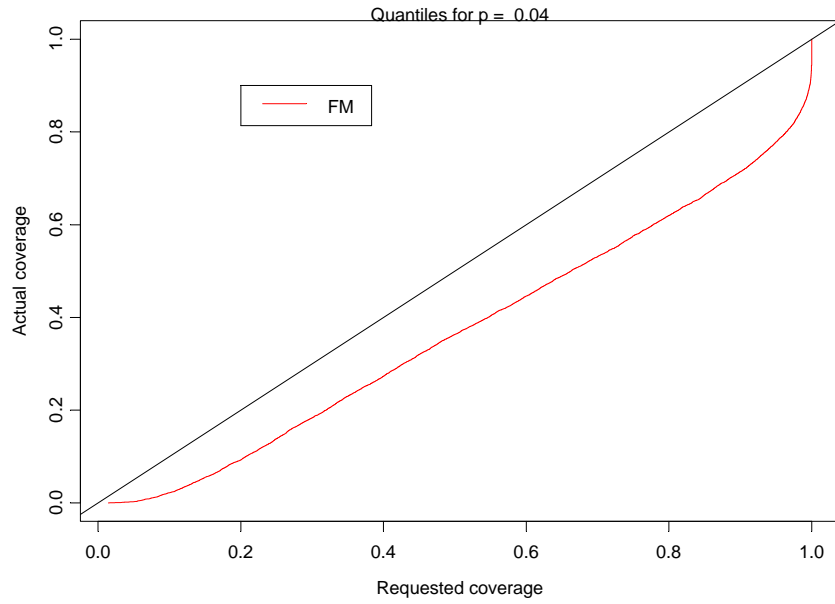
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 3$



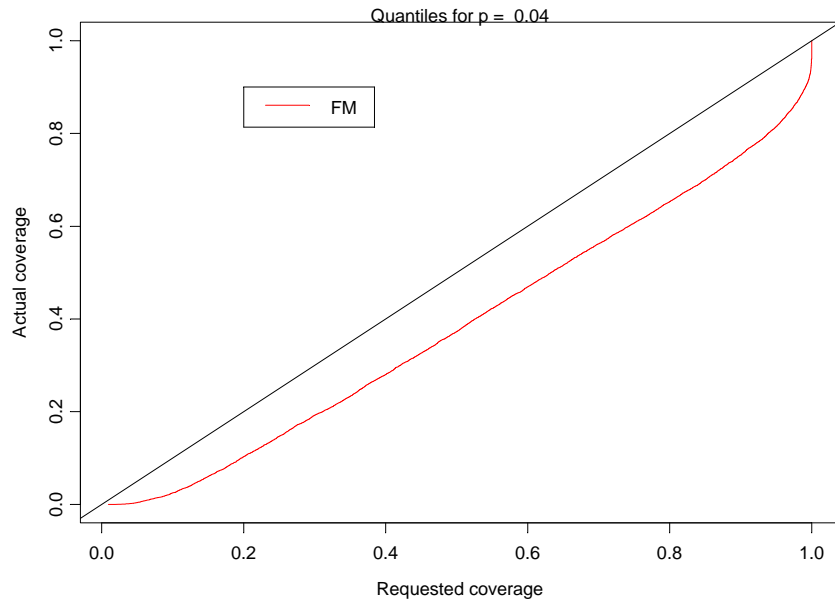
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 5$



Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 7$

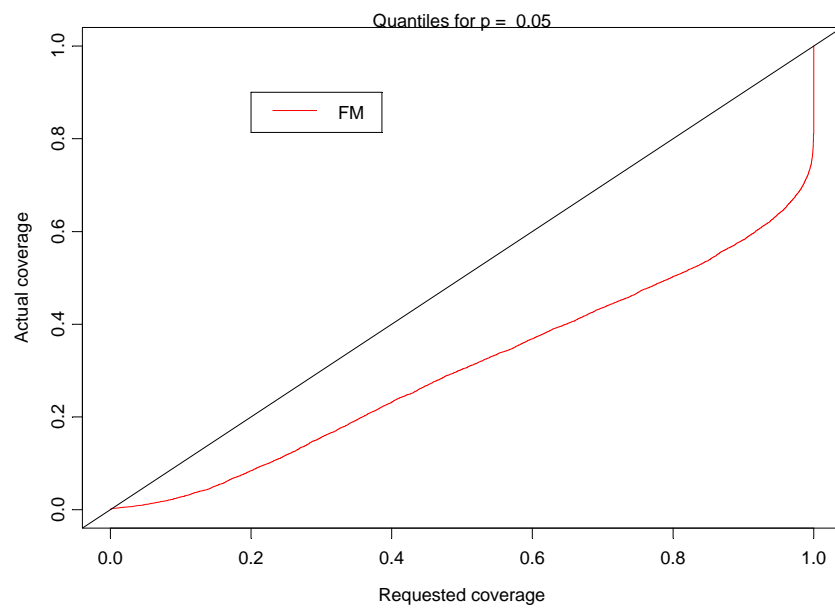


Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 9$

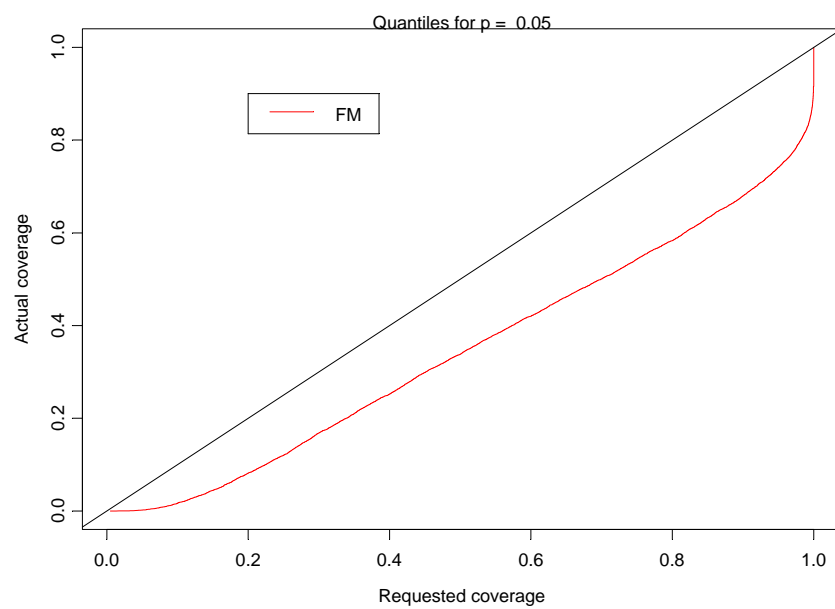


I. RELIABILITY = 95%, SAMPLE SIZE = 10

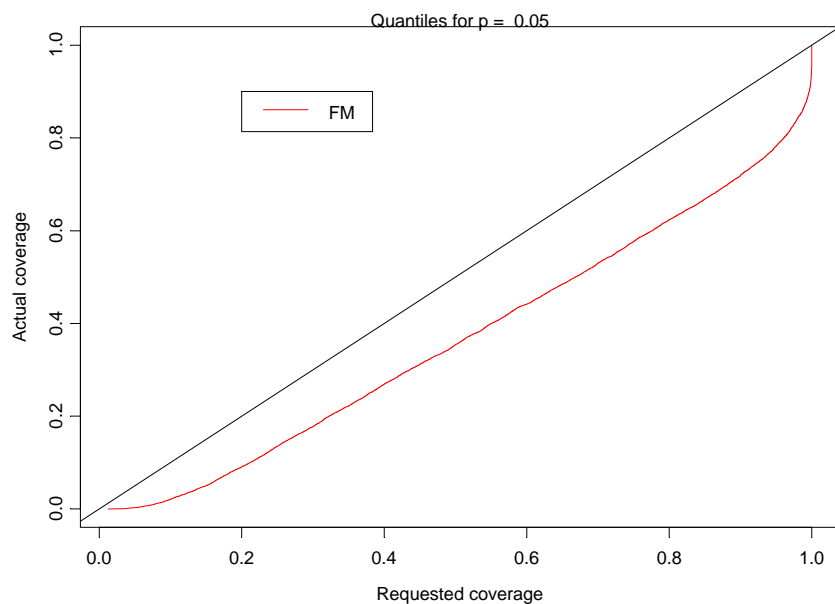
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 3$



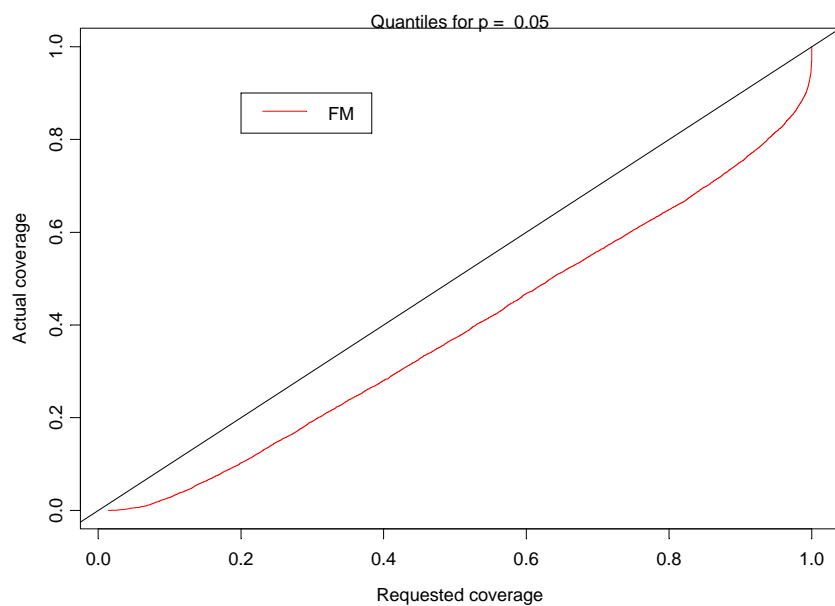
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 5$



Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 7$

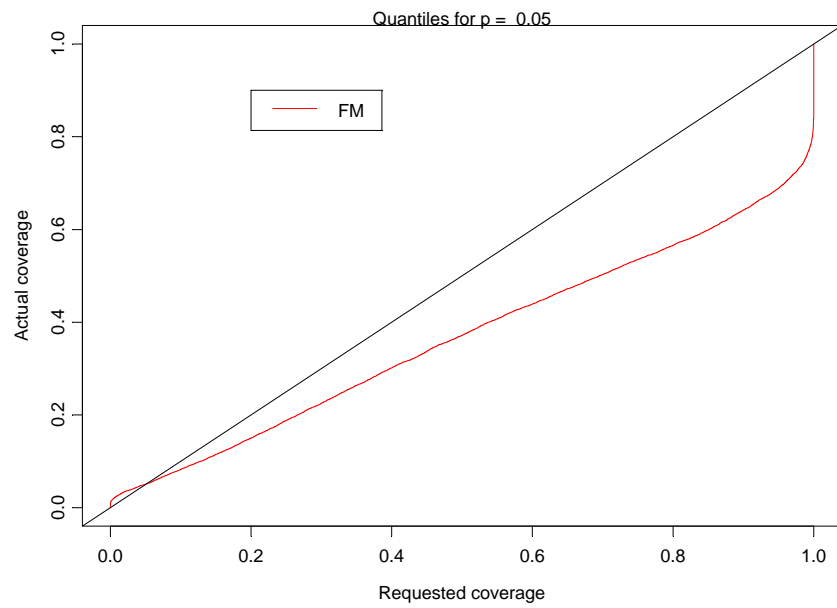


Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 9$

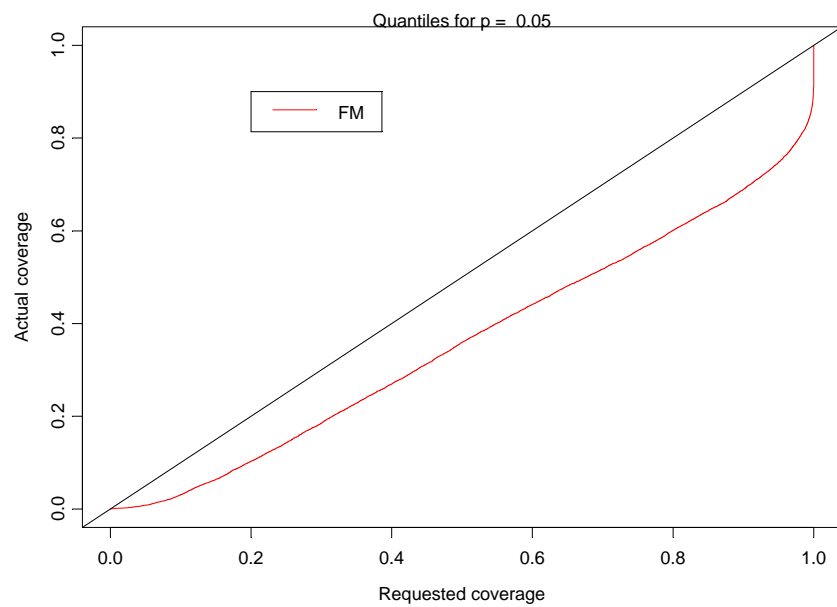


J. RELIABILITY = 95%, SAMPLE SIZE = 20

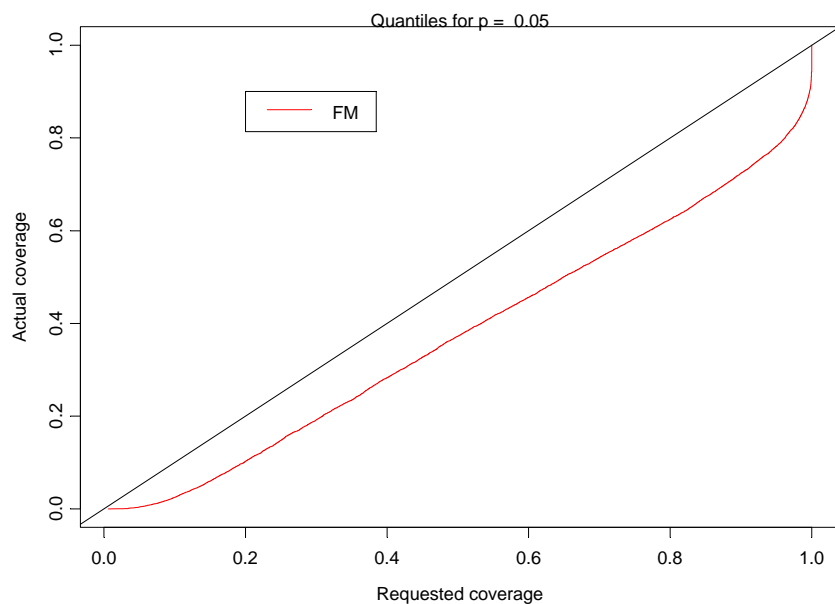
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 3$



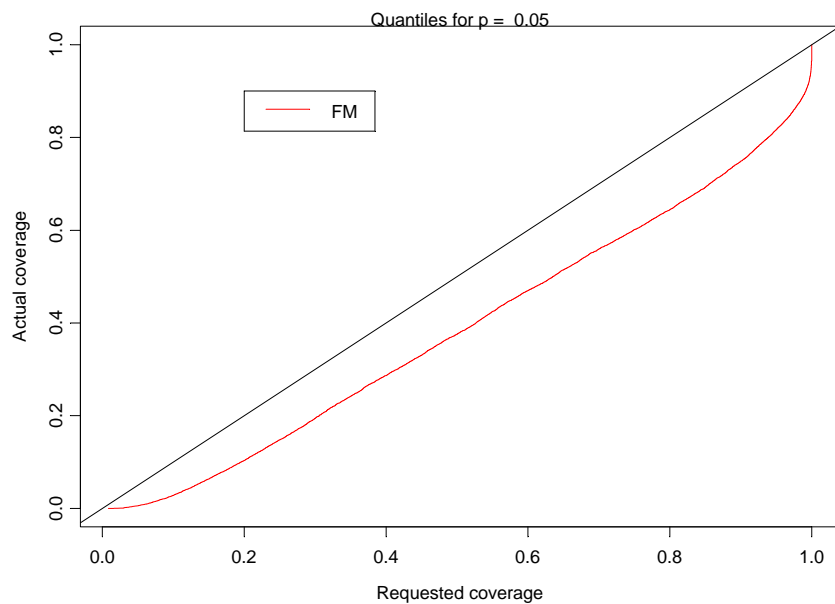
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 5$



Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 7$

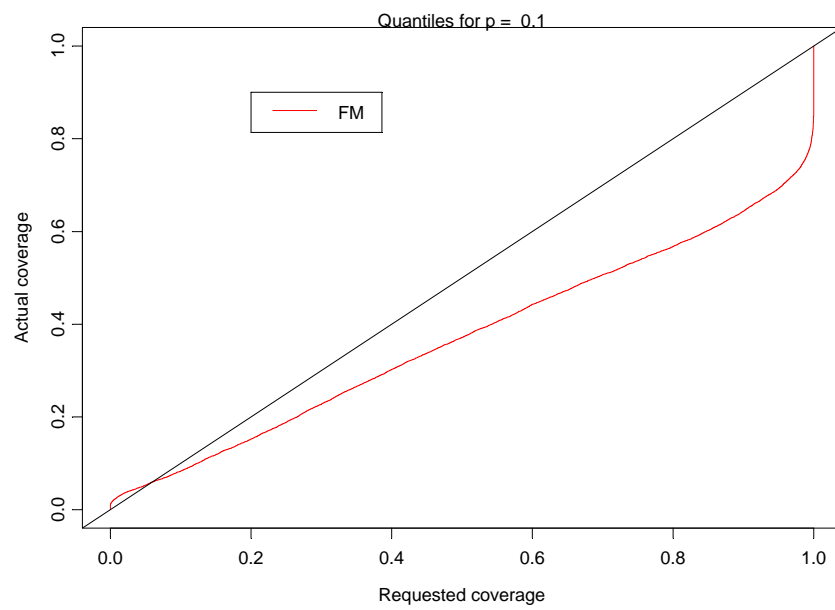


Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 9$

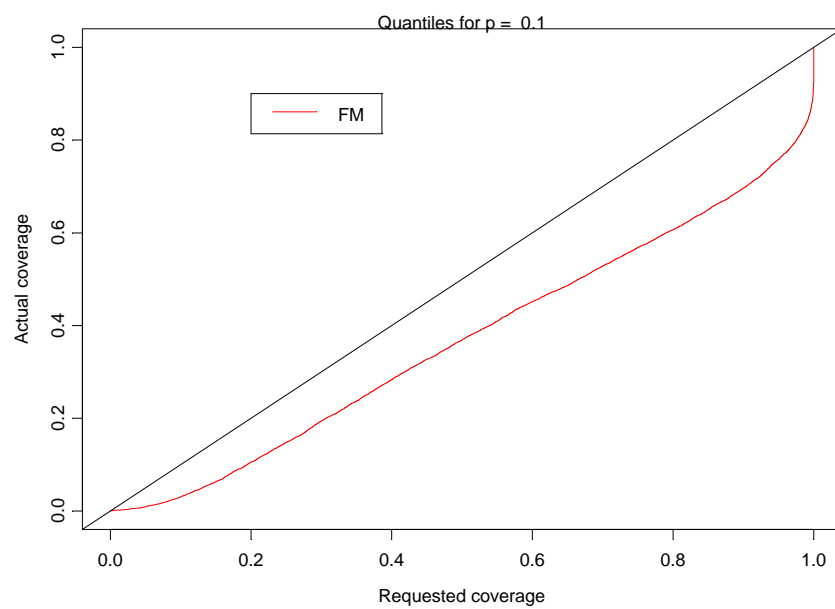


K. RELIABILITY = 90%, SAMPLE SIZE = 10

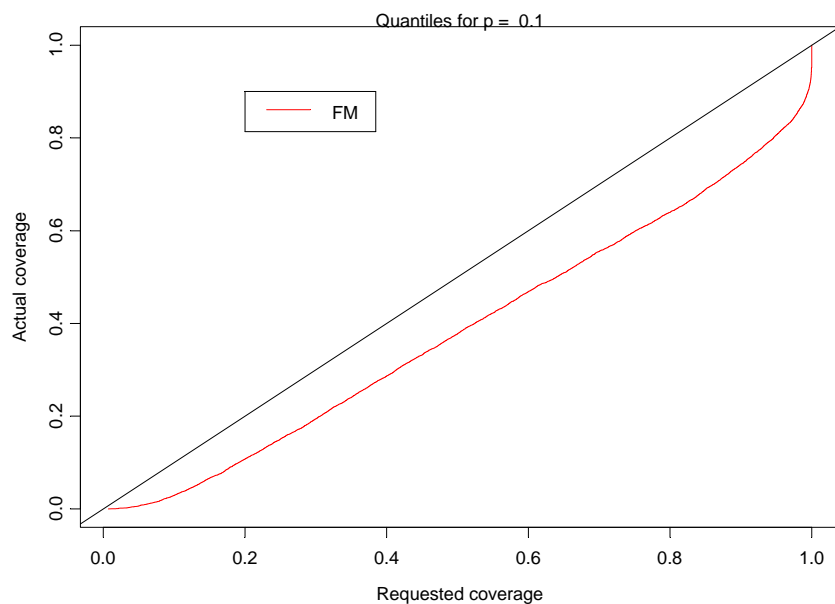
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 3$



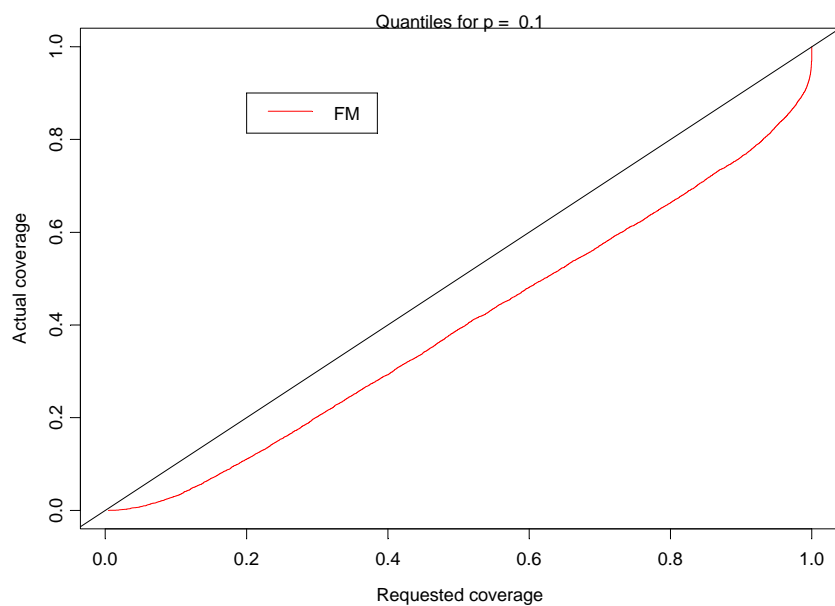
Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 5$



Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 7$

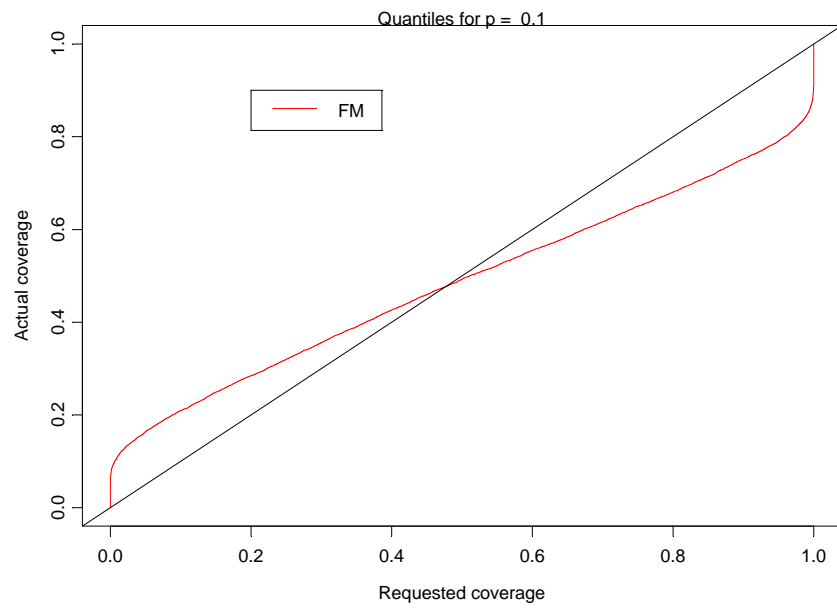


Actual vs. Requested coverage for $N = 10000$, $n = 10$, and $f = 9$

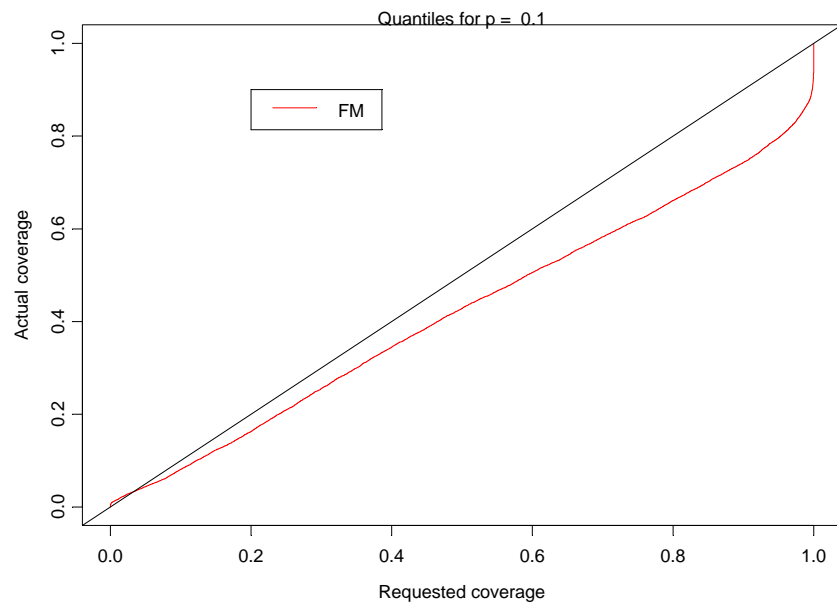


L. RELIABILITY = 90%, SAMPLE SIZE = 20

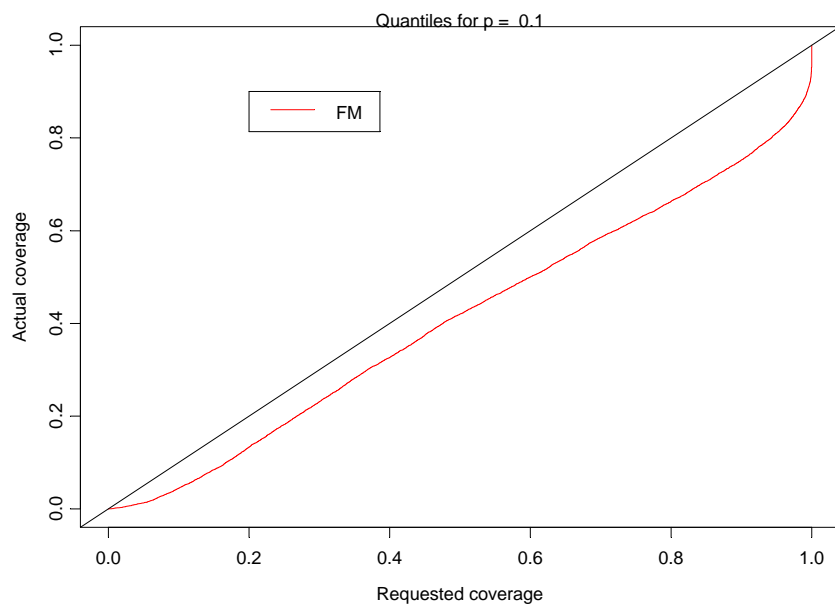
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 3$



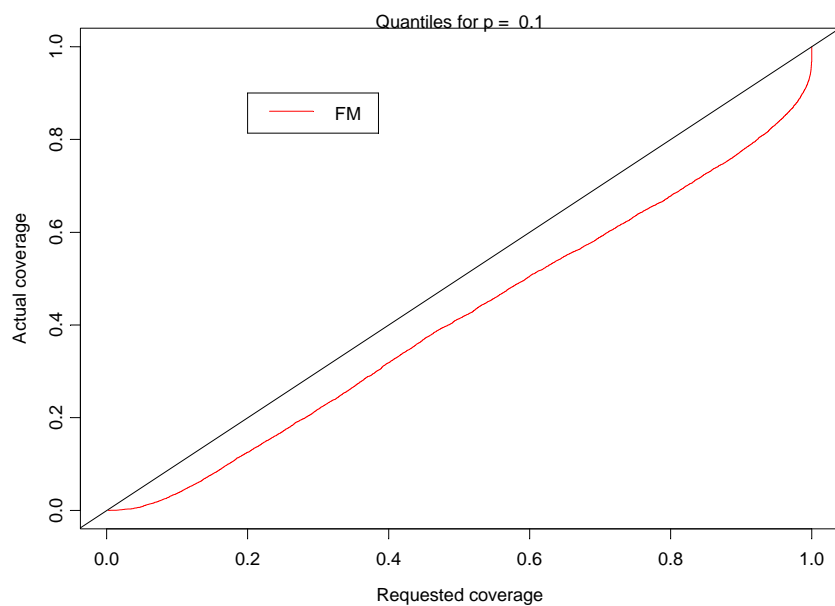
Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 5$



Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 7$



Actual vs. Requested coverage for $N = 10000$, $n = 20$, and $f = 9$



LIST OF REFERENCES

1. Nelson, W., Accelerated Testing: Statistical Models, Test Plans, and Data Analysis, John Wiley & Sons, Inc., 1990.
2. SPLIDA, William Q. Meeker,
<http://www.public.iastate.edu/~wqmeeker/Splida2000.html>. Last accessed March 02.
3. Meeker, W. Q. and Escobar, L. A., Statistical Methods for Reliability Data, John Wiley & Sons, Inc., 1998.
4. Nelson, W., Applied Life Data Analysis, John Wiley & Sons, Inc., 1982.
5. Devore, J. L., Probability and Statistics, Brooks/Cole Publishing, 2000.
6. Jeng, S. L. and Meeker, W. Q., “Comparisons of Approximate Confidence Interval Procedures for Type 1 Censored Data, Technometrics, Vol. 42, No. 2, May 2000.
7. S-Plus, Insightful Corporation, 2002, <http://www.insightful.com/splus>. Last accessed March 02.
8. NIST/SEMATECH e-Handbook of Statistical Methods, 2003,
<http://www.itl.nist.gov/div898/handbook/>. Last accessed Mar 02.
9. Anderson-Cook, C. M., An In-Class Demonstration to Help Students Understand Confidence Intervals, *Journal of Statistics Education* v.7, n.3, 1999,
<http://www.amstat.org/publications/jse/secure/v7n3/anderson-cook.cfm>. Last accessed March 02.
10. Lawless, J. F., Statistical Models and Methods for Lifetime Data, Wiley & Sons, Inc., 1982.
11. Aeronautical Research Laboratory Technical Report ARL 71-077, Confidence and Tolerance Bounds and a New Goodness of Fit Test for the Two Parameter Weibull or Extreme Value Distribution with Tables for Censored Samples of Size 3(1)25, by Mann, Fertig, and Scheur, 1971.
12. Weibull ++ 6, Life Data Analysis Software, ReliaSoft Corporation, 2003.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, VA
2. Dudley Knox Library
Naval Postgraduate School
Monterey, CA
3. Professor David H. Olwell
Operations Research Department
Monterey, CA
4. Professor Lyn Whitaker
Operations Research Department
Monterey, CA
5. Mr. Peter Burke
Edwards AFB, CA
6. Mr. Oscar L. Alvarado
Edwards AFB, CA
7. Mr. Michael A. Hartley
Edwards AFB, CA
8. Mr. Pantelis Vassiliou
ReliaSoft Corporation
Tucson, AZ